GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cggtgcaaatcgcccgatgg.....gcgtgccgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	33.4	33.1	434	10	BE122146	BE122146 894017F07
	2	33	32.7	632	10	BE582676	BE582676 9-10D-MY
	3	32.6	32.3	921	9	AL574395	AL574395 AL574395
c	4	32.2	31.9	1114	14	BQ050843	BQ050843 AGENCOURT
	5	32	31.7	476	9	AA024701	AA024701 ze79c04.s
c	6	31.8	31.5	421	9	AU163491	AU163491 AU163491
c	7	31.8	31.5	448	10	AV644360	AV644360 AV644360
c	8	31.8	31.5	452	9	AU094715	AU094715 AU094715
c	9	31.8	31.5	495	10	AV640754	AV640754 AV640754
c	10	31.8	31.5	497	10	AV628507	AV628507 AV628507
c	11	31.8	31.5	498	10	AV628086	AV628086 AV628086
c	12	31.8	31.5	503	10	AV638132	AV638132 AV638132
c	13	31.8	31.5	507	10	AV631846	AV631846 AV631846
c	14	31.8	31.5	511	10	AV391885	AV391885 AV391885
c	15	31.8	31.5	517	10	AV635168	AV635168 AV635168
c	16	31.8	31.5	534	10	AV621642	AV621642 AV621642
c	17	31.8	31.5	536	10	AV396567	AV396567 AV396567
c	18	31.8	31.5	635	13	BI874098	BI874098 963113D08
c	19	31.8	31.5	656	13	BI724994	BI724994 1031076D1
c	20	31.8	31.5	666	14	BQ810877	BQ810877 1030019F0
c	21	31.8	31.5	677	12	BG860320	BG860320 1024068G0
c	22	31.6	31.3	1300	14	BM919318	BM919318 AGENCOURT
	23	31.4	31.1	311	9	AA853014	AA853014 NHTBCae02
	24	31.4	31.1	427	9	AA668638	AA668638 ag45h11.s
	25	31.4	31.1	431	12	BF111987	BF111987 7137c06.x
	26	31.4	31.1	451	12	BF111254	BF111254 7174b03.x
	27	31.4	31.1	474	12	BG492131	BG492131 602535273
	28	31.4	31.1	512	9	AA516134	AA516134 ni07h04.s
	29	31.4	31.1	535	9	AA522923	AA522923 ni41f12.s
	30	31.4	31.1	585	12	BF435325	BF435325 nab38a10.
c	31	31.4	31.1	1630	10	BE421606	BE421606 HWM011cG.
	32	30.6	30.3	462	9	AL822466	AL822466 AL822466
	33	30.6	30.3	499	10	BE497906	BE497906 WHE0958_H
	34	30.6	30.3	718	14	BQ841046	BQ841046 WHE4206_D
	35	30.6	30.3	770	14	BQ841544	BQ841544 WHE4212_D
	36	30.6	30.3	847	9	AL569731	AL569731 AL569731
	37	30.4	30.1	272	12	BG057161	BG057161 7f74b08.x
	38	30.4	30.1	336	13	BM662318	BM662318 UI-E-CI1-
	39	30.4	30.1	427	12	BG223323	BG223323 naf79b01.
	40	30.4	30.1	456	9	AA705536	AA705536 zj92e01.s
	41	30.4	30.1	513	10	AW168364	AW168364 xi80g12.x
	42	30.4	30.1	629	14	BM971132	BM971132 UI-CF-EC1
	43	30.4	30.1	629	14	BQ183984	BQ183984 UI-H-EU0-
c	44	30.2	29.9	260	10	AV640218	AV640218 AV640218
c	45	30.2	29.9	435	13	BM259173	BM259173 952010C10

ALIGNMENTS

RESULT 1
BE122146/c
LOCUS BE122146 434 bp mRNA linear EST 13-JUN-2000
DEFINITION 894017F07.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BE122146
VERSION BE122146.1 GI:8514251
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 434)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Elizabeth H. Harris
DCMB Box 91000
Duke University
Durham, NC 27708-1000, USA
Tel: 919 613 8164
Fax: 919 613 8177
Email: chlamy@duke.edu.
FEATURES Location/Qualifiers
source 1. .434
/organism="Chlamydomonas reinhardtii"

```
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT       67 a    120 c    163 g     84 t
ORIGIN

  Query Match             33.1%;  Score 33.4;  DB 10;  Length 434;
  Best Local Similarity   60.4%;  Pred. No. 28;
  Matches    55;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| |||     ||||   ||   |  |
Db     364 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGCCCACGCACCAGACC 305

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||| ||
Db     304 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 274


RESULT 2
BE582676
LOCUS       BE582676                632 bp    mRNA    linear   EST 16-AUG-2000
DEFINITION  9-10D-MY PsojaeMY Phytophthora sojae cDNA, mRNA sequence.
ACCESSION   BE582676
VERSION     BE582676.1  GI:9833624
KEYWORDS    EST.
SOURCE      Phytophthora sojae.
  ORGANISM  Phytophthora sojae
            Eukaryota; stramenopiles; Comycetes; Pythiales; Pythiaceae;
            Phytophthora.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.
  TITLE     Comparative analysis of expressed sequences in Phytophthora sojae
  JOURNAL   Plant Physiol. 123 (1), 243-254 (2000)
  MEDLINE   20267956
COMMENT     Contact: Gijzen M
            Agriculture and Agri-Food Canada
            1391 Sandford Street, London, Ontario, Canada N5V 4T3
            Tel: 519 457 1470
            Fax: 519 457 3997
            Email: gijzenm@em.agr.ca.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Phytophthora sojae"
                     /strain="race 2, strain P6497"
                     /db_xref="taxon:67593"
                     /clone_lib="PsojaeMY"
                     /dev_stage="mycelium"
                     /lab_host="E. coli strain XLOLR"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
                     cDNA library was constructed from polyA+ enriched mRNA
                     from mycelium grown in liquid medium for 3 weeks at 25 C
                     in the dark  in a liquid synthetic medium of 2.4 g sucrose
                     , 0.27 g asparagine,  10 mg cholesterol, 10 mg ascorbate,
                     and 2 mg thiamine, per liter,  plus salts and minerals.
                     Complementary DNA was synthesized from  mRNA using an
                     XhoI-poly(dT) linker-primer. EcoRI adapters were ligated
                     to the blunt-ended cDNA fragments and the products were
                     digested with XhoI for directional cloning into lambda ZAP
                     Express vector. This lambda library was amplified once
                     using E. coli host strain XL1 Blue MRF'. Inserts were then
                     subcloned by mass excision using ExAssist helper phage for
                     conversion into phagemid vector pBK-CMV in E. coli host
                     strain XLOLR. Sequenced using T3 primer: 5' ATT AAC CCT
                     CAC TAA AGG GA 3'."
BASE COUNT      157 a    172 c    192 g    105 t      6 others
ORIGIN

  Query Match             32.7%;  Score 33;  DB 10;  Length 632;
  Best Local Similarity   63.0%;  Pred. No. 38;
  Matches    51;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       4 TGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGG 63
           | || |||||     | | || |||||||| ||||||||   ||||    |||   ||
Db       9 TCCAGATCGCCAAGGCGCTGGGCGCCGAGGTGATCGCCACGTGCAGCCACCGCAACGTGG 68

Qy      64 CAGCGGTTCAGGCCATGGGCG 84
               |||  || |  ||||||
Db      69 AGCTGGTCAAGTCGCTGGGCG 89


RESULT 3
AL574395
LOCUS       AL574395                921 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL574395 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI058YP20 3
            prime, mRNA sequence.
ACCESSION   AL574395
VERSION     AL574395.1  GI:12934563
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YP20"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      151 a    265 c    344 g    149 t      12 others
ORIGIN

  Query Match             32.3%;  Score 32.6;  DB 9;  Length 921;
  Best Local Similarity   59.6%;  Pred. No. 52;
  Matches    53;  Conservative    1;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      13 CCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTC 72
           |||  |||   |||:  |||  || |    |   |    ||||| |||| |||||||| |
Db      87 CCCCGTGGGRGGGGSTGGAGAGGTGCAGGCCCAGGGCGTCGTCCGAGTGAGCAGCGGTGC 146

Qy      73 AGGCCATGGGCGTGCCGCGCACGCACATC 101
             |||| || | |  ||| || |||  | |
Db     147 TGGCCGTGTGGGACCCGGGCCCGCTTAGC 175
```

```
RESULT 4
BQ050843/c
LOCUS       BQ050843                1114 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6941136 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5785616
            5', mRNA sequence.
ACCESSION   BQ050843
VERSION     BQ050843.1  GI:19810171
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1114)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12873  row: a  column: 09
            High quality sequence stop: 475.
FEATURES             Location/Qualifiers
     source          1. .1114
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5785616"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
BASE COUNT      312 a    325 c    270 g    206 t      1 others
ORIGIN

  Query Match             31.9%;  Score 32.2;  DB 14;  Length 1114;
  Best Local Similarity   62.8%;  Pred. No. 68;
  Matches   49;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      19 GGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCA 78
           |||    || || | ||||   |||| | ||| |||| ||   ||| |||    ||||
Db    1083 GGAGTCTGGGCGGGTTGTTACTCACGCCCAGCGCGTCGAATGCGGCGGCGAGCNAGGCGG 1024

Qy      79 TGGGCGTGCCGCGCACGC 96
           ||  |||||||||   ||
Db    1023 TGCTCGTGCCGCGACAGC 1006


RESULT 5
AA024701
LOCUS       AA024701                476 bp    mRNA    linear   EST 01-FEB-1997
DEFINITION  ze79c04.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
            IMAGE:365190 3' similar to PIR:B40572 B40572 collagen alpha 2(VI)
            chain precursor, long form - human ;, mRNA sequence.
ACCESSION   AA024701
VERSION     AA024701.1  GI:1489614
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 593   Std Error: 0.00
            Seq primer: -40M13 fwd. from Amersham
            High quality sequence stop: 365.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Homo sapiens"
                     /db_xref="GDB:1281894"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:365190"
                     /clone_lib="Soares_fetal_heart_NbHH19W"
                     /sex="unknown"
                     /dev_stage="19 weeks"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by
                     M.Fatima Bonaldo. This library was constructed from the
                     same fetus as the fetal lung library, Soares fetal lung
                     NbHL19W."
BASE COUNT       74 a    126 c    191 g     79 t      6 others
ORIGIN

  Query Match             31.7%;  Score 32;  DB 9;  Length 476;
  Best Local Similarity   59.6%;  Pred. No. 68;
  Matches   53;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      13 CCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTC 72
           ||||  |||  | |   |||  || |    |   |    ||||| |||| |||||||| |
Db    157 CCCGTGGGANGGAGCTGGAGAGGTGCAGGCCCAGGGCGTCGTCCGAGTGAGCAGCGGTGC 216

Qy      73 AGGCCATGGGCGTGCCGCGCACGCACATC 101
            |||| || | |  ||| || |||  | |
Db    217 TGGCCGTGTGGGACCCGGGCCCGCTTAGC 245


RESULT 6
AU163491/c
LOCUS       AU163491                421 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU163491 Rice cDNA from young root Oryza sativa (japonica
            cultivar-group) cDNA clone R10853, mRNA sequence.
ACCESSION   AU163491
VERSION     AU163491.1  GI:11118303
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 421)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from young root (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
```

```
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES              Location/Qualifiers
     source           1. .421
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="R10853"
                      /clone_lib="Rice cDNA from young root"
                      /tissue_type="young root"
BASE COUNT       69 a    136 c    138 g     73 t      5 others
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 9;  Length 421;
  Best Local Similarity   64.0%;  Pred. No. 75;
  Matches   48;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      20 GATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCAT 79
           |||  | |||  |||| ||| | |||||  | |||| |||| | | ||| | |  |||
Db     383 GATCAGCGCCCCGGTGCTCGTCCCGGCGATCACGTCGAAGTAGTCCGCGATCCTCGCCTC 324

Qy      80 GGGCGTGCCGCGCAC 94
           |||  | |  || |
Db     323 CGGCCCGTCCAGCTC 309


RESULT 7
AV644360/c
LOCUS       AV644360                 448 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV644360 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HCL087a05_r 5', mRNA sequence.
ACCESSION   AV644360
VERSION     AV644360.1  GI:10787688
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 448)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES              Location/Qualifiers
     source           1. .448
                      /organism="Chlamydomonas reinhardtii"
                      /strain="C9"
                      /db_xref="taxon:3055"
                      /clone="HCL087a05_r"
                      /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                      /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                      XhoI; The cDNA library was constructed from cells cultured
                      in a medium with bubbling air containing 5% carbon
                      dioxide"
BASE COUNT       75 a    124 c    161 g     88 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 448;
  Best Local Similarity   59.3%;  Pred. No. 76;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| |||| |||||| |||| |||| | ||  ||  | |
Db     247 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 188

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     187 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 157


RESULT 8
AU094715/c
LOCUS       AU094715                 452 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU094715 Rice panicle (longer than 10cm) Oryza sativa (japonica
            cultivar-group) cDNA clone E20707, mRNA sequence.
ACCESSION   AU094715
VERSION     AU094715.1  GI:8857397
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 452)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle (longer than 10cm) (2000)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES              Location/Qualifiers
     source           1. .452
                      /organism="Oryza sativa (japonica cultivar-group)"
                      /cultivar="Nipponbare"
                      /db_xref="taxon:39947"
                      /clone="E20707"
                      /clone_lib="Rice panicle (longer than 10cm)"
                      /tissue_type="panicle"
                      /dev_stage="panicle (longer than 10cm)"
                      /note="Organ: panicle; Rice cDNA from panicle at ripening
                      stage"
BASE COUNT       78 a    150 c    139 g     85 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 9;  Length 452;
  Best Local Similarity   64.0%;  Pred. No. 76;
  Matches   48;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      20 GATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCAT 79
           |||  | |||  |||| ||| | |||||| | |||| |||| | | ||| | |  |||
Db     420 GATCAGCGCCCCGGTGCTCGTCCCGGCGATCACGTCGAAGTAGTCCGCGATCCTCGCCTC 361

Qy      80 GGGCGTGCCGCGCAC 94
           |||  | |  || |
Db     360 CGGCCCGTCCAGCTC 346


RESULT 9
AV640754/c
LOCUS       AV640754                 495 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV640754 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HCL021h03_r 5', mRNA sequence.
ACCESSION   AV640754
VERSION     AV640754.1  GI:10784082
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
```

```
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HCL021h03_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       77 a    145 c    175 g     98 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 495;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     367 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 308

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     307 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 277


RESULT 10
AV628507/c
LOCUS       AV628507                497 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV628507 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL042h12_r 5', mRNA sequence.
ACCESSION   AV628507
VERSION     AV628507.1  GI:10791141
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCL042h12_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       77 a    148 c    174 g     98 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 497;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     370 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 311

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     310 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 280


RESULT 11
AV628086/c
LOCUS       AV628086                498 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV628086 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL036e01_r 5', mRNA sequence.
ACCESSION   AV628086
VERSION     AV628086.1  GI:10790720
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 498)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .498
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCL036e01_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       78 a    148 c    174 g     98 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 498;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     370 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 311

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     310 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 280


RESULT 12
AV638132/c
LOCUS       AV638132                503 bp    mRNA    linear   EST 15-DEC-2000
```

```
DEFINITION  AV638132 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HC082d03_r 5', mRNA sequence.
ACCESSION   AV638132
VERSION     AV638132.1  GI:10781452
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 503)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .503
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HC082d03_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       79 a    147 c    178 g     99 t
ORIGIN

  Query Match           31.5%;  Score 31.8;  DB 10;  Length 503;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     367 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 308

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     307 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 277


RESULT 13
AV631846/c
LOCUS       AV631846                 507 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV631846 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL100g04_r 5', mRNA sequence.
ACCESSION   AV631846
VERSION     AV631846.1  GI:10794480
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 507)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .507
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCL100g04_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       83 a    149 c    176 g     99 t
ORIGIN

  Query Match           31.5%;  Score 31.8;  DB 10;  Length 507;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     370 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 311

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
           | | || | | |||||| |  || |||||||
Db     310 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 280


RESULT 14
AV391885/c
LOCUS       AV391885                 511 bp    mRNA    linear   EST 23-APR-2002
DEFINITION  AV391885 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
            cDNA clone CM081e12_r 5', mRNA sequence.
ACCESSION   AV391885
VERSION     AV391885.1  GI:6546101
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
  TITLE     A large scale structural analysis of cDNAs in a unicellular green
            alga, Chlamydomonas reinhardtii. I. Generation of 3433
            non-redundant expressed sequence tags
  JOURNAL   DNA Res. 6 (6), 369-373 (1999)
  MEDLINE   20152988
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .511
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="CM081e12_r"
                     /clone_lib="Chlamydomonas reinhardtii C9"
                     /dev_stage="photoautotrophic growth"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       81 a    149 c    181 g    100 t
ORIGIN

  Query Match           31.5%;  Score 31.8;  DB 10;  Length 511;
  Best Local Similarity   59.3%;  Pred. No. 77;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
           ||  | || |   | ||| ||||    ||||||| ||||     | ||   ||   |  |
Db     357 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 298

Qy      65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
```

```
          | | || | | |||||| |  || |||||||
Db    297 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 267


RESULT 15
AV635168/c
LOCUS       AV635168                517 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV635168 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
            cDNA clone HC042g08_r 5', mRNA sequence.
ACCESSION   AV635168
VERSION     AV635168.1  GI:10778488
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="HC042g08_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a medium with bubbling air containing 5% carbon
                     dioxide"
BASE COUNT       81 a    151 c    184 g    101 t
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 517;
  Best Local Similarity   59.3%;  Pred. No. 78;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
          ||  | || |   | ||| ||||  . ||||||| ||||     | ||   ||   |  |
Db    357 GCCCACCGGCACCTTGATGGGGGAGCCGGTGTTCACCACCTTCTGGCCACGCACCAGACC 298

Qy     65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACG 95
          | | || | | |||||| |  || |||||||
Db    297 ATCAGTGCTGTCCATGGCCACGCAGCGCACG 267


Search completed: November  5, 2002, 16:18:23
Job time : 1118.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cggtgcaaatcgcccgatgg.....gcgtgccgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	37	36.6	11220	10	US-09-861-289-32	Sequence 32, Appl
2	37	36.6	36778	10	US-09-861-289-5	Sequence 5, Appli
c 3	29.8	29.5	1642	10	US-09-765-231A-25	Sequence 25, Appl
c 4	28.4	28.1	2640	10	US-09-917-800A-1726	Sequence 1726, Ap
5	28.2	27.9	13058	10	US-09-764-846-303	Sequence 303, App
c 6	27.6	27.3	2494	10	US-09-833-381-1984	Sequence 1984, Ap
c 7	27.4	27.1	4908	9	US-10-001-887-33	Sequence 33, Appl
c 8	26.8	26.5	4661	10	US-09-285-385C-3	Sequence 3, Appli
c 9	26	25.7	342	10	US-09-983-965-3684	Sequence 3684, Ap
c 10	26	25.7	349	10	US-09-960-352-71	Sequence 71, Appl
c 11	26	25.7	372	10	US-09-983-965-2663	Sequence 2663, Ap
c 12	26	25.7	873	10	US-09-891-718-3	Sequence 3, Appli
c 13	26	25.7	1716	10	US-09-919-781-3	Sequence 3, Appli
c 14	26	25.7	1917	10	US-09-919-781-1	Sequence 1, Appli
c 15	26	25.7	2424	10	US-09-742-312-1	Sequence 1, Appli
c 16	26	25.7	2724	10	US-09-891-718-1	Sequence 1, Appli
c 17	26	25.7	2787	10	US-09-939-408A-16	Sequence 16, Appl
c 18	26	25.7	147309	10	US-09-742-312-3	Sequence 3, Appli
19	25.8	25.5	6000	10	US-09-764-860-768	Sequence 768, App
c 20	25.8	25.5	43058	10	US-09-954-456-292	Sequence 292, App
c 21	25.8	25.5	43058	10	US-09-954-456-529	Sequence 529, App
c 22	25.8	25.5	43058	10	US-09-880-107-3950	Sequence 3950, Ap
c 23	25.4	25.1	373	10	US-09-864-761-21426	Sequence 21426, A
24	25.4	25.1	659	10	US-09-815-242-3978	Sequence 3978, Ap
c 25	25.2	25.0	1352	10	US-09-896-856-2	Sequence 2, Appli
c 26	25.2	25.0	2430	10	US-09-861-289-23	Sequence 23, Appl
27	25.2	25.0	3089	10	US-09-989-722-46	Sequence 46, Appl
28	25.2	25.0	3089	10	US-09-989-723-46	Sequence 46, Appl
29	25.2	25.0	3089	10	US-09-989-279-46	Sequence 46, Appl
30	25.2	25.0	3089	10	US-09-989-727-46	Sequence 46, Appl
31	25.2	25.0	3089	10	US-09-989-731-46	Sequence 46, Appl
32	25.2	25.0	3089	10	US-09-989-732-46	Sequence 46, Appl
33	25.2	25.0	3089	10	US-09-991-073-46	Sequence 46, Appl
34	25.2	25.0	3089	10	US-09-990-442-46	Sequence 46, Appl
35	25.2	25.0	3089	10	US-09-991-163-46	Sequence 46, Appl
36	25.2	25.0	3089	10	US-09-993-604-46	Sequence 46, Appl
37	25.2	25.0	3089	10	US-09-990-456-46	Sequence 46, Appl
38	25.2	25.0	3089	10	US-09-989-721-46	Sequence 46, Appl
39	25.2	25.0	3089	12	US-10-052-586-61	Sequence 61, Appl
c 40	25.2	25.0	13613	10	US-09-861-289-3	Sequence 3, Appli
41	25	24.8	534	10	US-09-738-973-265	Sequence 265, App
42	25	24.8	894	10	US-09-815-242-4029	Sequence 4029, Ap
c 43	25	24.8	1464	10	US-09-815-242-7915	Sequence 7915, Ap
44	25	24.8	1503	10	US-09-815-242-7680	Sequence 7680, Ap
45	25	24.8	1590	12	US-10-042-417-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 36.6%; Score 37; DB 10; Length 11220;
Best Local Similarity 60.4%; Pred. No. 0.014;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           | |||||  |||||||        || |  |||||   || |||||||||  |     ||||
Db    9314 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 9373

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
             |||  || ' | |  ||  | ||| ||     | ||||||||
Db    9374 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 9414
```

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             36.6%;  Score 37;  DB 10;  Length 36778;
  Best Local Similarity   60.4%;  Pred. No. 0.014;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db   25001 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 25060

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  ||  | ||| ||     | ||||||||
Db   25061 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 25101


RESULT 3
US-09-765-231A-25/c
; Sequence 25, Application US/09765231A
; Patent No. US20020119452A1
; GENERAL INFORMATION:
;  APPLICANT: Searle/Monsanto
;  APPLICANT:  Phippard, Deborah
;  APPLICANT:  Vasanthakamur, Geetha
;  APPLICANT:  Dotson, Stanton
;  APPLICANT:  Ma, Xiao-Jun
;  TITLE OF INVENTION: Osteoarthritis tissue-derived nucleic acids, polypeptides,
;  TITLE OF INVENTION:  vectors, and cells
;  FILE REFERENCE: SO-3221 PR
;  CURRENT APPLICATION NUMBER: US/09/765,231A
;  CURRENT FILING DATE:  2001-01-18
;  NUMBER OF SEQ ID NOS: 82
; SEQ ID NO 25
;   LENGTH: 1642
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-765-231A-25

  Query Match             29.5%;  Score 29.8;  DB 10;  Length 1642;
  Best Local Similarity   58.4%;  Pred. No. 1.5;
  Matches   52;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      13 CCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTC 72
           |||  |||   | |   |||  || |    |   |    ||||| |||| |||||||| |
Db    1476 CCCCGTGGGAGGAGCTGGAGAGGTGCAGGCCCAGGGCGTCGTCCGAGTGAGCAGCGGTGC 1417

Qy      73 AGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||| || | |  ||| || |||  | |
Db    1416 TGGCCGTGTGGGACCCGGGCCCGCTTAGC 1388


RESULT 4
US-09-917-800A-1726/c
; Sequence 1726, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1726
;   LENGTH: 2640
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_024163
US-09-917-800A-1726

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 2640;
  Best Local Similarity   56.4%;  Pred. No. 4;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       8 AATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGC 67
             |||| || ||| |    |||  |||     | ||    | |||   ||||| | |   |
Db     278 AATCTCCAGATAGTGGCGGGTGGAGTCAAACTCCGTCTCCAGCTTCTCCAACTTGTGCGT 219

Qy      68 GGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||| || | || |||   || ||| | ||
Db     218 GGTGTAGGACAGGCGCTTGCGCAGCTCGCCCTTC 185


RESULT 5
US-09-764-846-303
; Sequence 303, Application US/09764846
; Patent No. US20020102638A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ12
;  CURRENT APPLICATION NUMBER: US/09/764,846
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 303
;   LENGTH: 13058
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-846-303

  Query Match             27.9%;  Score 28.2;  DB 10;  Length 13058;
  Best Local Similarity   61.6%;  Pred. No. 4.8;
  Matches   45;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       4 TGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGG 63
           ||||||    | | |||   |||| | ||||  |  ||| | ||  |||||||   || |
```

```
Db   2502 TGCAAAGACGCGGCTGGGCTGGGGACTAGGTAGTGCCCAGGCCGGCCCCGTCCCTCTGAG 2561

Qy     64 CAGCGGTTCAGGC 76
          |||    || |||
Db   2562 CAGGAACTCTGGC 2574


RESULT 6
US-09-833-381-1984/c
; Sequence 1984, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: No. US20020132090A1el Nucleic Acid and Protein Homologs
;  FILE REFERENCE: 5800-119
;  CURRENT APPLICATION NUMBER: US/09/833,381
;  CURRENT FILING DATE:  2001-04-11
;  PRIOR APPLICATION NUMBER: 09/516,448
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 2050
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1984
;   LENGTH: 2494
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-381-1984

  Query Match            27.3%;  Score 27.6;  DB 10;  Length 2494;
  Best Local Similarity  63.6%;  Pred. No. 6.7;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     25 GGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCATGGGCG 84
          ||   | |||  | ||   ||||  ||||    |||||| |||||  |||||  | ||||
Db   2201 GGCAGGTGGTCCTGGCGGGGGCGGCCCCGCAGGAGTGGGGAGCGGCGCAGGCGGTCGGCG 2142

Qy     85 TGCCGC 90
          | | ||
Db   2141 TCCTGC 2136


RESULT 7
US-10-001-887-33/c
; Sequence 33, Application US/10001887
; Patent No. US20020155464A1
; GENERAL INFORMATION:
;  APPLICANT: Salceda, Susana
;  APPLICANT:  Macina, Roberto
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Cafferkey, Robert
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and Pr
;  FILE REFERENCE: DEX-0269
;  CURRENT APPLICATION NUMBER: US/10/001,887
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/249,998
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: 60/252,563
;  PRIOR FILING DATE: 2000-11-22
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
;   LENGTH: 4908
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-10-001-887-33

  Query Match            27.1%;  Score 27.4;  DB 9;  Length 4908;
  Best Local Similarity  59.7%;  Pred. No. 7.9;
  Matches   46;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy     15 CGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAG 74
          ||| || |  |||| | || || |||  | | | |  || |   | || | | ||| |

Db    211 CGAGGGCTCCGGGCAGGGGCGTCCGCGGCTGGGCGGGCGGCGTTGGGGTCCGGGGTCCGC 152

Qy     75 GCCATGGGCGTGCCGCG 91
          |||   || || |||||
Db    151 GCCGGCGGGGTTCCGCG 135


RESULT 8
US-09-285-385C-3/c
; Sequence 3, Application US/09285385C
; Patent No. US20020150946A1
; GENERAL INFORMATION:
;  APPLICANT: Greenspan, Daniel S.
;  APPLICANT:  Scott, Ian C.
;  APPLICANT:  Thomas, Christina L.
;  TITLE OF INVENTION: MAMMALIAN TOLLOID-LIKE GENE AND PROTEIN
;  FILE REFERENCE: 960296.96111
;  CURRENT APPLICATION NUMBER: US/09/285,385C
;  CURRENT FILING DATE:  1999-04-02
;  PRIOR APPLICATION NUMBER: 60/111873
;  PRIOR FILING DATE: 1998-12-11
;  PRIOR APPLICATION NUMBER: 60/080550
;  PRIOR FILING DATE: 1998-04-03
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 4661
;   TYPE: DNA
;   ORGANISM: mouse
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (191)..(3229)
US-09-285-385C-3

  Query Match            26.5%;  Score 26.8;  DB 10;  Length 4661;
  Best Local Similarity  55.3%;  Pred. No. 12;
  Matches   52;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      5 GCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGC 64
          ||| |     | | |||| |||||| || |||  | | |  ||  |   |||| |||| |
Db   2734 GCAGAAGCGGCCAAGGATGGGGGCCAAGCTGTCTGTCCCATCGTACAGTTCCAGGTGGTC 2675

Qy     65 AGCGGTTCAGGCCATGGGCGTGCCGCGCACGCAC 98
          |  ||  ||  ||  | ||    |   | || ||
Db   2674 ATAGGCACATTCCTGGTGCTGCTCAATCTCGAAC 2641


RESULT 9
US-09-983-965-3684/c
; Sequence 3684, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 3684
;   LENGTH: 342
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 43-LIB3058-043-Q1-K1-C12
US-09-983-965-3684
```

```
  Query Match             25.7%;  Score 26;  DB 10;  Length 342;
  Best Local Similarity   59.5%;  Pred. No. 18;
  Matches   44;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      19 GGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCA 78
           ||   |||| | || ||  |  |  ||   ||| |  | ||  ||  |||| |||| |
Db     182 GGGGCGGGGACAAGCTGGGCTGCCAGGGCCGCCGGGTCCAGGCGGGCGCGGAGCAGGGCG 123

Qy      79 TGGGCGTGCCGCGC 92
           ||||| || ||||
Db     122 CGGGCGAGCTGCGC 109


RESULT 10
US-09-960-352-71/c
; Sequence 71, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 71
;   LENGTH: 349
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 01-LIB3057-002-Q1-K1-A9
US-09-960-352-71

  Query Match             25.7%;  Score 26;  DB 10;  Length 349;
  Best Local Similarity   59.5%;  Pred. No. 18;
  Matches   44;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      19 GGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCA 78
           ||   |||| | || ||  |  |  ||   ||| |  | ||  ||  |||| |||| |
Db     259 GGGGCGGGGACAAGCTGGGCTGCCAGGGCCGCCGGGTCCAGGCGGGCGCGGAGCAGGGCG 200

Qy      79 TGGGCGTGCCGCGC 92
           ||||| || ||||
Db     199 CGGGCGAGCTGCGC 186


RESULT 11
US-09-983-965-2663/c
; Sequence 2663, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2663
;   LENGTH: 372
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 10-LIB3058-043-Q1-K1-C5
US-09-983-965-2663

  Query Match             25.7%;  Score 26;  DB 10;  Length 372;
  Best Local Similarity   59.5%;  Pred. No. 18;
  Matches   44;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      19 GGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCA 78
           ||   |||| | || ||  |  |  ||   ||| |  | ||  ||  |||| |||| |
Db     149 GGGGCGGGGACAAGCTGGGCTGCCAGGGCCGCCGGGTCCAGGCGGGTGCGGAGCAGGGCG 90

Qy      79 TGGGCGTGCCGCGC 92
           ||||| || ||||
Db      89 CGGGCGAGCTGCGC 76


RESULT 12
US-09-891-718-3/c
; Sequence 3, Application US/09891718
; Patent No. US20020127671A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A. J.
;  APPLICANT:  Silos-Santiago, Inmaculada
;  TITLE OF INVENTION: 52927, A NOVEL HUMAN CALCIUM CHANNEL AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: MNI-168
;  CURRENT APPLICATION NUMBER: US/09/891,718
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,101
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 873
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(873)
US-09-891-718-3

  Query Match             25.7%;  Score 26;  DB 10;  Length 873;
  Best Local Similarity   62.1%;  Pred. No. 19;
  Matches   41;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      22 TAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCATGG 81
           ||||    || ||  ||| | ||  |  || || | || |||| | |  ||||||  |||
Db     500 TAGGCATAGAAGTCGTCGTCGCGCGGCTCCTGTTCCAGCGGGCTGAGCGTCAGGCGCTGG 441

Qy      82 GCGTGC 87
           | |||
Db     440 CCCTGC 435


RESULT 13
US-09-919-781-3/c
; Sequence 3, Application US/09919781
; Patent No. US20020123094A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  TITLE OF INVENTION: 57250, A NOVEL HUMAN SUGAR TRANSPORTER
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREOF
;  FILE REFERENCE: MNI-181
;  CURRENT APPLICATION NUMBER: US/09/919,781
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: USSN 60/221,769
;  PRIOR FILING DATE: 2000-07-31
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1716
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1716)
US-09-919-781-3

  Query Match             25.7%;  Score 26;  DB 10;  Length 1716;
  Best Local Similarity   62.1%;  Pred. No. 19;
  Matches   41;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      25 GGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCATGGGCG 84
           ||   | |||  | ||   ||||  ||||    ||||||  ||||  |||||  | ||||
Db    1664 GGCAGGTGGTCCTGGCGGGGGCGGCCCCGCAGGAGTGGGGGGCGGCGCAGGCGGTCGGCG 1605

Qy      85 TGCCGC 90
           | | ||
Db    1604 TCCTGC 1599


RESULT 14
US-09-919-781-1/c
; Sequence 1, Application US/09919781
; Patent No. US20020123094A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A.J.
;  TITLE OF INVENTION: 57250, A NOVEL HUMAN SUGAR TRANSPORTER
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREOF
;  FILE REFERENCE: MNI-181
;  CURRENT APPLICATION NUMBER: US/09/919,781
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: USSN 60/221,769
;  PRIOR FILING DATE: 2000-07-31
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1917
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13)...(1728)
US-09-919-781-1

  Query Match             25.7%;  Score 26;  DB 10;  Length 1917;
  Best Local Similarity   62.1%;  Pred. No. 19;
  Matches   41;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      25 GGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCATGGGCG 84
           ||   | |||  | ||   ||||  ||||    ||||||  ||||  |||||  | ||||
Db    1676 GGCAGGTGGTCCTGGCGGGGGCGGCCCCGCAGGAGTGGGGGGCGGCGCAGGCGGTCGGCG 1617

Qy      85 TGCCGC 90
           | | ||
Db    1616 TCCTGC 1611


RESULT 15
US-09-742-312-1/c
; Sequence 1, Application US/09742312
; Patent No. US20020045166A1
; GENERAL INFORMATION:
;  APPLICANT: CHANDRAMOULISARAN, Ishwar et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000838
;  CURRENT APPLICATION NUMBER: US/09/742,312
;  CURRENT FILING DATE:  2000-12-22
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 2424
;   TYPE: DNA
;   ORGANISM: Human
US-09-742-312-1

  Query Match             25.7%;  Score 26;  DB 10;  Length 2424;
  Best Local Similarity   62.1%;  Pred. No. 20;
  Matches   41;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      22 TAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGGTTCAGGCCATGG 81
           ||||    || ||  ||| | ||  |  || || | || |||| | |  ||||||  |||
Db    461 TAGGCATAGAAGTCGTCGTCGCGCAGCTCCTGTTCCAGCGGGCTGAGCGTCAGGCGCTGG 402

Qy      82 GCGTGC 87
           | |||
Db    401 CCCTGC 396


Search completed: November  5, 2002, 23:01:55
Job time : 94.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cggtgcaaatcgcccgatgg.....gcgtgccgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	49.8	49.3	44377	2	US-08-804-227C-7	Sequence 7, Appli
	10	49.8	49.3	44377	2	US-08-804-198-1	Sequence 1, Appli
c	11	39.2	38.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	12	37	36.6	11220	4	US-09-105-537-32	Sequence 32, Appl
	13	37	36.6	36778	4	US-09-105-537-5	Sequence 5, Appli
	14	37	36.6	38506	3	US-09-320-878-19	Sequence 19, Appl
	15	36	35.6	20235	1	US-07-642-734C-3	Sequence 3, Appli
	16	36	35.6	20235	3	US-08-439-009A-3	Sequence 3, Appli
	17	35.6	35.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
	18	35.6	35.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
	19	35.6	35.2	80161	4	US-09-370-700-1	Sequence 1, Appli
	20	35.4	35.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
	21	35.4	35.0	28958	1	US-08-456-837-6	Sequence 6, Appli
	22	35.4	35.0	28958	1	US-08-457-342-6	Sequence 6, Appli
	23	35.4	35.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
	24	35.4	35.0	28958	1	US-08-458-076A-6	Sequence 6, Appli
	25	35.4	35.0	28958	1	US-08-764-233A-4	Sequence 4, Appli
	26	35.4	35.0	28958	1	US-08-457-335A-6	Sequence 6, Appli
	27	35.4	35.0	28958	1	US-08-729-214-6	Sequence 6, Appli
	28	35.4	35.0	28958	3	US-09-028-934-6	Sequence 6, Appli
	29	35.4	35.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
	30	35.4	35.0	50937	4	US-09-428-517-1	Sequence 1, Appli
	31	32.6	32.3	352	4	US-08-818-112-23	Sequence 23, Appl
	32	32.6	32.3	352	4	US-08-818-111-23	Sequence 23, Appl
	33	32.6	32.3	352	4	US-09-056-556-23	Sequence 23, Appl
	34	32.6	32.3	352	4	US-09-072-596-23	Sequence 23, Appl
c	35	28.4	28.1	1038	4	US-09-403-768-3	Sequence 3, Appli
c	36	28.2	27.9	487	3	US-08-983-045-1	Sequence 1, Appli
c	37	28.2	27.9	1683	3	US-08-983-045-3	Sequence 3, Appli
c	38	28.2	27.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
c	39	28.2	27.9	44377	2	US-08-804-198-1	Sequence 1, Appli
	40	27.8	27.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	41	27.8	27.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	42	27.4	27.1	9515	1	US-08-920-812-13	Sequence 13, Appl
c	43	27.4	27.1	9515	1	US-08-920-827-13	Sequence 13, Appl
c	44	27.4	27.1	9515	1	US-08-921-177-13	Sequence 13, Appl
c	45	27.4	27.1	9515	1	US-08-362-577C-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 9.1e-20;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 68750
;    TYPE: DNA
;    ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 68750
;    TYPE: DNA
;    ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;    LENGTH: 71989
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29800 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 29859

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29860 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 29900


RESULT 9
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  317-276-2459
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  44377 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  350..14002
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  14046..20036
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  20110..31284
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31329..36071
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             49.3%;  Score 49.8;  DB 2;  Length 44377;
  Best Local Similarity   68.3%;  Pred. No. 1.3e-05;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || || |||||| | ||||| ||||||||   |||||
Db   29396 CGGTGCAGATCGCCCGCCACCTCGGCGCGGAGGTGCTGGCCACCGCGAGCCCCGGCAAGT 29455

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  |||||||||| |  |   |   ||| ||
Db   29456 GGGACGCCCTGCGCGCCATGGGCATCACCGACGACCACCTC 29496


RESULT 10
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             49.3%;  Score 49.8;  DB 2;  Length 44377;
  Best Local Similarity   68.3%;  Pred. No. 1.3e-05;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || || |||||| | ||||| ||||||||   |||||
Db   29396 CGGTGCAGATCGCCCGCCACCTCGGCGCGGAGGTGCTGGCCACCGCGAGCCCCGGCAAGT 29455

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  |||||||||| |  |   |   ||| ||
Db   29456 GGGACGCCCTGCGCGCCATGGGCATCACCGACGACCACCTC 29496


RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             38.8%;  Score 39.2;  DB 4;  Length 4403765;
  Best Local Similarity   62.0%;  Pred. No. 0.014;
  Matches   62;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           |||||||  | || ||     | |||   |||||||||| || ||||||  |   ||||
Db 2304759 CGGTGCAGCTGGCTCGGCATCTGGGGTTGGAGGTGTTCGCGACCGCGAGCAAGGGTAAGT 2304700

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACAT 100
           |||   |  | |  ||||||||| |      |  |||||
```

```
Db 2304699 GGGACACCTTGCGCGCCATGGGCTTTGACGACGACCACAT 2304660


RESULT 12
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             36.6%;  Score 37;  DB 4;  Length 11220;
  Best Local Similarity   60.4%;  Pred. No. 0.04;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db   9314 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 9373

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||  ||  | |  ||  | ||| ||     | ||||||||
Db   9374 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 9414


RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             36.6%;  Score 37;  DB 4;  Length 36778;
  Best Local Similarity   60.4%;  Pred. No. 0.043;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db  25001 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 25060

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||  ||  | |  ||  | ||| ||     | ||||||||
Db  25061 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 25101


RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             36.6%;  Score 37;  DB 3;  Length 38506;
  Best Local Similarity   60.4%;  Pred. No. 0.043;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db  23143 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 23202

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||  ||  | |  ||  | ||| ||     | ||||||||
Db  23203 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 23243


RESULT 15
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
```

```
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Danckers, Andreas M
;       REGISTRATION NUMBER:  32652
;       REFERENCE/DOCKET NUMBER:  4952.US.01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  708-937-9396
;       TELEFAX:  708-938-2623
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  20235 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharopolyspora erythraea
;       STRAIN:  NRRL 238
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  19..10722
;       OTHER INFORMATION:  /codon_start= 19
;       OTHER INFORMATION:  /function= 'gene eryA"
;       OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;       OTHER INFORMATION:  6-deoxyerythronolide B"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19..4470
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= 'approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3

  Query Match             35.6%;  Score 36;  DB 1;  Length 20235;
```

Best Local Similarity 62.0%; Pred. No. 0.078;
Matches 57; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy      10 TCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGG 69
           | |||||  |    |||||  |||||||| |||||  ||  |||||  |||||   |   ||
Db    8948 TGGCCCGCCGAGCGGGGGCGGAGGTGTTGGCCACCGCCGGCCCGGCCAAGCACGGGACGC 9007

Qy      70 TTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           | | |||  | ||  |      |  |||||||
Db    9008 TGCGGGCGCTCGGTCTCGACGACGAGCACATC 9039
```

Search completed: November 5, 2002, 16:03:55
Job time : 1239.74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cggtgcaaatcgcccgatgg.....gcgtgccgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	101	100.0	71989	21	AAA29349	Sorangium cellulos
	3	49.8	49.3	44377	18	AAT78508	Platenolide syntha
	4	49.8	49.3	44377	18	AAT80414	Platenolide syntha
	5	49.8	49.3	125401	22	AAD17186	Streptomyces nours
	6	46.6	46.1	65140	22	AAD17184	Streptomyces nours
	7	40.8	40.4	47981	22	AAF30757	Micromonospora meg
c	8	39.2	38.8	4403765	22	AAI99683	Mycobacterium tube
	9	37	36.6	11220	21	AAZ87298	S. venezuelae macr
	10	37	36.6	36778	21	AAZ87318	S. venezuelae pik
	11	37	36.6	37948	21	AAZ87285	S. venezuelae pik
	12	37	36.6	38506	21	AAA75633	Nucleotide sequenc
	13	37	36.6	38506	21	AAZ56001	Recombinant cosmid
	14	37	36.6	77536	21	AAA14651	Nucleotide sequenc
	15	36	35.6	29879	14	AAQ46806	eryA region of S.
	16	35.6	35.2	4851	22	AAH52062	Mycobacterium tube
	17	35.6	35.2	6459	22	AAF88336	S. spinosa DNA fra
	18	35.6	35.2	43280	18	AAT80413	Tylactone synthase
	19	35.6	35.2	50000	22	AAF88313	S. spinosa DNA fra
	20	35.6	35.2	50000	22	AAF88316	S. spinosa DNA fra
	21	35.6	35.2	80161	20	AAZ21501	DNA fragment of Sa
	22	35.4	35.0	28598	17	AAT06769	Sorangium cellulos
	23	35.4	35.0	28958	18	AAT89956	Sorangium cellulos
	24	35.4	35.0	28958	21	AAA75299	DNA sequence of So
	25	35.4	35.0	49377	19	AAV05287	The soraphen biosy
	26	35.4	35.0	50937	21	AAA09469	Streptococcus olea
c	27	32.8	32.5	30690	21	AAA92301	S. avermitilis ave
c	28	32.8	32.5	30690	22	AAH79277	Streptomyces averm
	29	32.6	32.3	352	18	AAT91483	Mycobacterium tube
	30	32.6	32.3	352	18	AAT91420	Mycobacterium tube
	31	32.6	32.3	352	19	AAV64469	M. tuberculosis im
	32	32.6	32.3	352	19	AAV44361	Mycobacterium tube
	33	32.6	32.3	352	20	AAZ19271	M. tuberculosis an
	34	32.6	32.3	352	20	AAZ19059	M. tuberculosis re
c	35	30.6	30.3	77536	21	AAA14651	Nucleotide sequenc
	36	30.2	29.9	1263	22	AAD08879	Mycobacterium szul
c	37	30	29.7	909	21	AAA11259	Burkholderia strai
	38	30	29.7	1365	24	ABQ90174	M. capsulatus gene
c	39	29.8	29.5	1642	22	AAH23095	Osteoarthritis tis
	40	29.2	28.9	1985	21	AAZ36421	DNA encoding a mai
c	41	29.2	28.9	9359	22	AAK73225	Human immune/haema
c	42	28.8	28.5	1656	22	AAF45121	Human TANGO 202 cD
	43	28.6	28.3	6375	23	ABL14967	Drosophila melanog
c	44	28.6	28.3	8792	23	ABL14966	Drosophila melanog
	45	28.6	28.3	124884	22	AAH74201	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes ER5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match           100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29800 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 29859

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           ||||||||||||||||||||||||||| ||||||||||
Db   29860 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 29900


RESULT 3
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match           49.3%;  Score 49.8;  DB 18;  Length 44377;
  Best Local Similarity   68.3%;  Pred. No. 0.00011;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || || |||||| | ||||| ||||||||   |||||
Db   29396 CGGTGCAGATCGCCCGCCACCTCGGCGCGGAGGTGCTGGCCACCGCGAGCCCCGGCAAGT 29455

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  ||||||||| |  |   |   ||| ||
Db   29456 GGGACGCCCTGCGCGCCATGGGCATCACCGACGACCACCTC 29496


RESULT 4
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
```

```
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match            49.3%;  Score 49.8;  DB 18;  Length 44377;
  Best Local Similarity  68.3%;  Pred. No. 0.00011;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || || |||||| | ||||| ||||||||   |||||
Db   29396 CGGTGCAGATCGCCCGCCACCTCGGCGCGGAGGTGCTGGCCACCGCGAGCCCCGGCAAGT 29455

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  |||||||||| |  |   |   ||| ||
Db   29456 GGGACGCCCTGCGCGCCATGGGCATCACCGACGACCACCTC 29496


RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
```

```
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             49.3%;  Score 49.8;  DB 22;  Length 125401;
  Best Local Similarity   68.3%;  Pred. No. 0.00011;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           || | || ||||||||     | || || ||||||||||||||||| ||   |  |||||
Db   39503 CGATCCAGATCGCCCGGCATGTCGGTGCCGAGGTGTTCGCCACGGCCAGTGAGGGCAAGT 39562

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  | | | |   || |||||||| |   |   ||||||
Db   39563 GGGACGTGCTGCGCTCCCTGGGCGTGGCCGACGACCACATC 39603


RESULT 6
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             46.1%;  Score 46.6;  DB 22;  Length 65140;
  Best Local Similarity   66.3%;  Pred. No. 0.0008;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           || | || ||||||||     | || ||||||||||||||||||||| ||   |  |||||
Db   31785 CGATCCAGATCGCCCGGCATGTCGGTGCCGAGGTGTTCGCCACGGCCAGTGAGGGCAAGT 31844

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  | | | |   |  | |||||| |   |   ||||||
Db   31845 GGGACGTGTTGCGCTCGCTCGGCGTGGCCGACGACCACATC 31885


RESULT 7
```

```
AAF30757
ID   AAF30757 standard; DNA; 47981 BP.
XX
AC   AAF30757;
XX
DT   21-JUN-2001  (first entry)
XX
DE   Micromonospora megalomicea megalomicin biosynthetic gene cluster.
XX
KW   Megalomicin; meg gene; polyketide synthase; antibiotic;
KW   motilide; antiparasitic; ds.
XX
OS   Micromonospora megalomicea subsp. nigra.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..144)
FT                   /*tag= a
FT                   /partial
FT                   /gene= "megBVI(megT)"T"
FT                   /product="TDP-4-keto-6-deoxyglucose-
FT                            2,3-dehydratase"
FT                   /note= "encodes AAB82201"
FT   CDS             928..2061
FT                   /*tag= b
FT                   /gene= "megDVI"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,4-isomerase"
FT                   /note= "encodes AAB82202"
FT   CDS             2072..3382
FT                   /*tag= c
FT                   /gene= "megDI"
FT                   /product= "TDP-megosamine glycosyltransferase"
FT                   /note= "eryCIII homologue; encodes AAB802203"
FT   CDS             3462..4634
FT                   /*tag= d
FT                   /gene= "megY"
FT                   /product= "mycarose O-acyltransferase"
FT                   /note= "encodes AAB82204"
FT   CDS             4651..5775
FT                   /*tag= e
FT                   /gene= "megDII"
FT                   /product= "TDP-3-keto-6-deoxyhexose
FT                             3-aminotransaminase"
FT                   /note= "eryCI, DnrJ homologue, encodes AAB82205"
FT   CDS             5822..6595
FT                   /*tag= f
FT                   /gene= "megDIII"
FT                   /product= "daunosaminyl-N,N-dimethyltransferase"
FT                   /note= "eryCVI homologue; encodes AAB82206"
FT   CDS             6592..7197
FT                   /*tag= g
FT                   /gene= "megDIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 3,5-epimerase"
FT                   /note= "eryBVIII, dnmU homologue, encodes AAB82207"
FT   CDS             7220..8206
FT                   /*tag= h
FT                   /gene= "megDV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "eryVIV, dnmV homolgoue; encodes AAB82208"
FT   CDS             complement (8228..9220)
FT                   /*tag= i
FT                   /gene= "megBII-1(megDVII)"
FT                   /product= "TDP-4-keto-6-deoxyhexose 2,3-reductase"
FT                   /note= "encodes AAB82209"
FT   CDS             complement (9226..10479)
FT                   /*tag= j
FT                   /gene= "megBV"
FT                   /product= "TDP-mycarose glycosyltransferase"
FT                   /note= "encodes AAB82210"
FT   CDS             complement (10483..11424)
FT                   /*tag= k
FT                   /gene= "megBIV"
FT                   /product= "TDP-4-keto-6-deoxyhexose 4-ketoreductase"
FT                   /note= "encodes AAB82211"
FT   CDS             12181..22821
FT                   /gene= "megA"
FT                   /*tag= l
FT                   /product= "megalomicin 6-deoxyerthronolide B
FT                             synthase 1"
FT                   /note= "polyketide synthase; encodes AAB82212"
FT   misc_feature    12505..13470
FT                   /*tag= m
FT                   /gene= "megA"
FT                   /function= "AT-L"
FT   misc_feature    13576..13791
FT                   /*tag= n
FT                   /gene= "megA"
FT                   /function= "ACP-L"
FT   misc_feature    13849..15126
FT                   /*tag= o
FT                   /gene= "megA"
FT                   /function= "KS1"
FT   misc_feature    15427..16476
FT                   /*tag= p
FT                   /gene= "megA"
FT                   /function= "AT1"
FT   misc_feature    17155..17694
FT                   /*tag= q
FT                   /gene= "megA"
FT                   /function= "KR1"
FT   misc_feature    17947..18207
FT                   /*tag= r
FT                   /gene= "megA"
FT                   /function= "ACP1"
FT   misc_feature    18268..19548
FT                   /*tag= s
FT                   /gene= "megA"
FT                   /function= "KS2"
FT   misc_feature    19876..20910
FT                   /*tag= t
FT                   /gene= "megA"
FT                   /function= "AT2"
FT   misc_feature    21517..22053
FT                   /*tag= u
FT                   /gene= "megA"
FT                   /function= "KR2"
FT   misc_feature    22318..22575
FT                   /*tag= v
FT                   /gene= "megA"
FT                   /function= "ACP2"
FT   CDS             22867..33555
FT                   /*tag= w
FT                   /gene= "megAII"
FT                   /product= "megalomicin 6-deoxyerythronolide B
FT                             synthase 2"
FT                   /note= "polyketide synthase, encodes AAB82213"
FT   misc_feature    22957..24237
FT                   /*tag= x
FT                   /gene= "megAII"
FT                   /function= "KS3"
FT   misc_feature    24544..25581
FT                   /*tag= y
FT                   /gene= "megAII"
FT                   /function= "AT3"
FT   misc_feature    26230..26733
FT                   /*tag= z
FT                   /gene= "megAII"
FT                   /function= "KR3 (inactive)"
FT   misc_feature    26998..27258
FT                   /*tag= aa
FT                   /gene= "megAII"
FT                   /function= "ACP3"
FT   misc_feature    27393..28590
FT                   /*tag= ab
FT                   /gene= "megAII"
FT                   /function= "KS4"
FT   misc_feature    28897..29931
```

```
FT                   /*tag=  ac
FT                   /gene= "megAII"
FT                   /function= "AT4"
FT   misc_feature    29953..30477
FT                   /*tag=  ad
FT                   /gene= "megAII"
FT                   /function= "DH4"
FT   misc_feature    31396..32244
FT                   /*tag=  ae
FT                   /gene= "megAII"
FT                   /function= "ER4"
FT   misc_feature    32257..32799
FT                   /*tag=  af
FT                   /gene= "megAII"
FT                   /function= "KR4"
FT   misc_feature    33052..33312
FT                   /*tag=  ag
FT                   /gene= "megAII"
FT                   /function= "ACP4"
FT   CDS             33666..43271
FT                   /*tag=  ah
FT                   /gene= "megAIII"
FT                   /product= "megalomicin 6-deoxyerythronolide B
FT                        synthase 3"
FT                   /note= "polyketide synthase; encodes AAB82214"
FT   misc_feature    22957..24237
FT                   /*tag=  ai
FT                   /gene= "megAIII"
FT                   /function= "KS5"
FT   misc_feature    24544..25581
FT                   /*tag=  aj
FT                   /gene= "megAIII"
FT                   /function= "AT5"
FT   misc_feature    26230..26733
FT                   /*tag=  ak
FT                   /gene= "megAIII"
FT                   /function= "KR5"
FT   misc_feature    26998..27258
FT                   /*tag=  al
FT                   /gene= "megAIII"
FT                   /function= "ACP5"
FT   misc_feature    27393..28590
FT                   /*tag=  am
FT                   /gene= "megAIII"
FT                   /function= "KS6"
FT   misc_feature    28897..29931
FT                   /*tag=  an
FT                   /gene= "megAIII"
FT                   /function= "AT6"
FT   misc_feature    29953..30477
FT                   /*tag=  ao
FT                   /gene= "megAIII"
FT                   /function= "KR6"
FT   misc_feature    31396..32244
FT                   /*tag=  ap

  Query Match             40.4%;  Score 40.8;  DB 22;  Length 47981;
  Best Local Similarity   65.2%;  Pred. No. 0.03;
  Matches   60;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      10 TCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGG 69
           | |||||  ||   ||||| ||||||||| ||||||| |||||| ||||     |  ||
Db   31772 TGGCCCGTCGGGCCGGGGCGGAGGTGTTCGCCACGGCCAGCCCGGCCAAACACCCGACGC 31831

Qy      70 TTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           | | |||  | ||| |      |   ||||||
Db   31832 TGCGGGCGCTCGGCCTCGACGACGACCACATC 31863


RESULT 8
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             38.8%;  Score 39.2;  DB 22;  Length 4403765;
  Best Local Similarity   62.0%;  Pred. No. 0.096;
  Matches   62;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           |||||||  | || ||     | |||   |||||||||||| || ||||||  |   ||||
Db 2304759 CGGTGCAGCTGGCTCGGCATCTGGGGTTGGAGGTGTTCGCGACCGCGAGCAAGGGTAAGT 2304700

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACAT 100
           |||   |  | |  |||||||| |      |   |||||
Db 2304699 GGGACACCTTGCGCGCCATGGGCTTTGACGACGACCACAT 2304660


RESULT 9
AAZ87298
ID   AAZ87298 standard; DNA; 11220 BP.
XX
AC   AAZ87298;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
```

```
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..111220
FT                   /*tag= a
FT                   /product= "PikAII"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77193.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 398-403; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

  Query Match             36.6%;  Score 37;  DB 21;  Length 11220;
  Best Local Similarity   60.4%;  Pred. No. 0.31;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           | |||||  |||||||       || |  |||||   || |||||||| |     ||||
Db    9314 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 9373

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  ||  | |  ||  | ||| ||     | ||||||||
Db    9374 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 9414
```

```
RESULT 10
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
```

CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC 15439, as given in figure 31.
XX
SQ Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

Query Match 36.6%; Score 37; DB 21; Length 36778;
Best Local Similarity 60.4%; Pred. No. 0.32;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
Db 25001 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 25060

Qy 61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
Db 25061 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 25101

RESULT 11
AAZ87285
ID AAZ87285 standard; DNA; 37948 BP.
XX
AC AAZ87285;
XX
DT 05-JUN-2000 (first entry)
XX
DE S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW chronic obstructive pulmonary disease; respiratory inflammation;
KW hypercholesterolaemia; crop protection agent; ds.
XX
OS Streptomyces venezuelae ATCC15439.
XX
PN WO200000620-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US14398.
XX
PR 26-JUN-1998; 98US-0105537.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Sherman DH, Liu H, Xue Y, Zhao L;
XX
DR WPI; 2000-160679/14.
DR P-PSDB; AAY77180.
XX
PT Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT synthesis of methymycin and pikromycin -
XX
PS Claim 13; Page 299-315; 438pp; English.
XX
CC The invention relates to an isolated and purified nucleic acid segment
CC comprising a desosamine biosynthetic gene cluster, a fragment or its
CC biologically active variant, where the nucleic acid sequence is not
CC derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC Streptomyces antibioticus. The invention also relates to a macrolide
CC biosynthetic gene cluster, or fragments thereof. The macrolide
CC biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC pikromycin, neomethymycin, narbomycin or a combination of these
CC compounds. Recombinant or augmented cells comprising the desosamine
CC and/or macrolide biosynthetic gene clusters are useful for the production
CC of biologically active macrolides. The macrolide biosynthetic proteins
CC are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC narbomycin. The alternative termination of polyketide synthesis may be
CC useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC monomers. The compounds produced by the recombinant host cells are useful
CC as biopolymers, e.g., in packaging or biomedical applications, to
CC engineer PHA monomer synthases or to prepare biologically active agents,
CC such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC chronic obstructive pulmonary disease as well as other diseases involving
CC respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC antibiotics which are active against a variety of organisms, e.g.,
CC bacteria, including multi-drug resistant pneumococci and other
CC respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC protection agents (e.g., fungicides or insecticides) via expression of
CC polyketides in plants. The present sequence represents the macrolide
CC biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX
SQ Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

Query Match 36.6%; Score 37; DB 21; Length 37948;
Best Local Similarity 60.4%; Pred. No. 0.32;
Matches 61; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
Db 26171 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 26230

Qy 61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
Db 26231 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 26271

RESULT 12
AAA75633
ID AAA75633 standard; DNA; 38506 BP.
XX
AC AAA75633;
XX
DT 22-JAN-2001 (first entry)
XX
DE Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW picromycin biosynthesis; ss.
XX
OS Streptomyces venezuelae.
XX
PN US6117659-A.
XX
PD 12-SEP-2000.
XX
PF 27-MAY-1999; 99US-0320878.
XX
PR 28-MAY-1998; 98US-0087080.
PR 22-SEP-1998; 98US-0100880.
PR 08-FEB-1999; 99US-0119139.
PR 20-MAY-1999; 99US-0134990.
PR 30-APR-1997; 97US-0846247.
PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and

```
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 15-32; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             36.6%;  Score 37;  DB 21;  Length 38506;
  Best Local Similarity   60.4%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db  23143 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 23202

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||  ||  | |  ||  | ||| ||     | ||||||||
Db  23203 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 23243


RESULT 13
AAZ56001
ID   AAZ56001 standard; DNA; 38506 BP.
XX
AC   AAZ56001;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
KW   antibiotic production; narbomycin; picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             70..13725
FT                   /*tag=  a
FT                   /product= PICAI
FT                   /note= "Narbonolide synthase subunit 1"
FT   CDS             13830..25049
FT                   /*tag=  b
FT                   /product= PICAII
FT                   /note= "Narbonolide synthase subunit 2"
FT   CDS             25133..29821
FT                   /*tag=  c
FT                   /product= PICAIII
FT                   /note= "Narbonolide synthase subunit 3"
FT   CDS             29924..33964
FT                   /*tag=  d
FT                   /product= PICAIV
FT                   /note= "Narbonolide synthase subunit 4"
FT   CDS             33961..34806
FT                   /*tag=  e
FT                   /product= PicB
FT                   /note= "Contains typeII thioesterase domain"
FT   CDS             34863..36011
FT                   /*tag=  f
FT                   /product= PICCII
FT                   /note= "4-keto-6-deoxyglucose isomerase"
FT   CDS             36159..37439
FT                   /*tag=  g
FT                   /product= PICCIII
FT                   /note= "Desosaminyl transferase"
FT   CDS             37529..38242
FT                   /*tag=  h
FT                   /product= PICCVI
FT                   /note= "3-amino dimethyltransferase"
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR   AAY67208, AAY67211.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 16-27; 98pp; English.
XX
CC   This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC   Streptomyces venezuelae DNA insert. The cosmid contains open reading
CC   frames which encode the various modules of the narbonolide polyketide
CC   synthase (PKS). The invention relates to recombinant DNA containing a
CC   coding sequence for a narbonolide PKS. Polyketides are compounds
CC   synthesised from 2-carbon units through a series of condensations and
CC   subsequent modifications. Modular PKSs are responsible for the production
CC   of many antibiotics including picromycin. The narbonolide PKS consists of
CC   a loading module, six extender modules, and two thioester domains. Four
CC   proteins make up the narbonolide PKS (PICAI, PICAII, PICAIII and PICAIV).
CC   PICAI includes the loading module and extender modules 1 and 2, PICAII
CC   includes extender modules 3 and 4, PICAIII includes extender module 5 and
CC   PICAIV includes extender module 6 and a type II thioesterase domain. The
CC   second type II thioesterase domain is found on the PICB protein. The
CC   nucleotide sequences encoding all of these proteins can be isolated in
CC   recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             36.6%;  Score 37;  DB 21;  Length 38506;
  Best Local Similarity   60.4%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          | |||||  |||||||       || |  |||||   || |||||||||| |     ||||
Db  23143 CCGTGCAGCTCGCCCGGCACTGGGGCGTGGAGGTCCACGGCACGGCGAGTCACGGGAAGT 23202

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||  ||  | |  ||  | ||| ||     | ||||||||
```

```
Db  23203 GGGACGCCCTGCGCGCGCTCGGCCTGGACGACGCGCACATC 23243


RESULT 14
AAA14651
ID   AAA14651 standard; DNA; 77536 BP.
XX
AC   AAA14651;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW   FK-506 binding protein; polyketide compound; transplant rejection;
KW   graft-versus-host disease; uveitis; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (412..1836)
FT                   /*tag= a
FT                   /note= "fkbW gene"
FT   CDS             complement (2020..3579)
FT                   /*tag= b
FT                   /note= "fkbV gene"
FT   CDS             3969..4496
FT                   /*tag= c
FT                   /note= "fkbR2 gene'
FT   CDS             complement (4595..5488)
FT                   /*tag= d
FT                   /note= "fkbR1 gene'
FT   CDS             5601..6818
FT                   /*tag= e
FT                   /note= "fkbE gene"
FT   CDS             6808..8052
FT                   /*tag= f
FT                   /note= "fkbF gene"
FT   CDS             8156..8824
FT                   /*tag= g
FT                   /note= "fkbG gene"
FT   CDS             complement (9122..9883)
FT                   /*tag= h
FT                   /note= "fkbH gene"
FT   CDS             complement (9894..10994)
FT                   /*tag= i
FT                   /note= "fkbI gene"
FT   CDS             complement (10987..11247)
FT                   /*tag= j
FT                   /note= "fkbJ gene"
FT   CDS             complement (11244..12092)
FT                   /*tag= k
FT                   /note= "fkbK gene"
FT   CDS             complement (12113..13150)
FT                   /*tag= l
FT                   /note= "fkbL gene"
FT   CDS             complement (13212..23988)
FT                   /*tag= m
FT                   /note= "fkbC gene"
FT   misc_feature    complement (13452..13662)
FT                   /*tag= n
FT                   /note= "ACP6"
FT   misc_feature    complement (13761..14394)
FT                   /*tag= o
FT                   /note= "KR6"
FT   misc_feature    complement (14517..15294)
FT                   /*tag= p
FT                   /note= "ER6"
FT   misc_feature    complement (15438..16587)
FT                   /*tag= q
FT                   /note= "dehydratase domain (DH) 6"
FT   misc_feature    complement (16587..17820)
FT                   /*tag= r
FT                   /note= "acyltransferase domain (AT) 6"
FT   misc_feature    complement (17820..19053)
FT                   /*tag= s
FT                   /note= "KS6"
FT   misc_feature    complement (19116..19326)
FT                   /*tag= t
FT                   /note= "ACP5"
FT   misc_feature    complement (19464..20097)
FT                   /*tag= u
FT                   /note= "KR5"
FT   misc_feature    complement (20241..21420)
FT                   /*tag= v
FT                   /note= "DH5"
FT   misc_feature    complement (21420..22653)
FT                   /*tag= w
FT                   /note= "AT5"
FT   misc_feature    complement (22653..23892)
FT                   /*tag= x
FT                   /note= "KS5"
FT   CDS             complement (23992..46573)
FT                   /*tag= y
FT                   /note= "fkbB gene"
FT   misc_feature    complement (24163..24373)
FT                   /*tag= z
FT                   /note= "ACP4"
FT   misc_feature    complement (24997..26146)
FT                   /*tag= aa
FT                   /note= "DH4 (inactive)"
FT   misc_feature    complement (26146..27430)
FT                   /*tag= ab
FT                   /note= "AT4"
FT   misc_feature    complement (27430..28684)
FT                   /*tag= ac
FT                   /note= "KS4"
FT   misc_feature    complement (28750..28960)
FT                   /*tag= ad
FT                   /note= "ACP3"
FT   misc_feature    complement (29092..29740)
FT                   /*tag= ae
FT                   /note= "KR3"
FT   misc_feature    complement (29869..31018)
FT                   /*tag= af
FT                   /note= "DH3 (inactive)"
FT   misc_feature    complement (31018..32185)
FT                   /*tag= ag
FT                   /note= "AT3"
FT   misc_feature    complement (32185..33439)
FT                   /*tag= ah
FT                   /note= "KS3"
FT   misc_feature    complement (33505..33715)
FT                   /*tag= ai
FT                   /note= "ACP2"
FT   misc_feature    complement (33823..34480)
FT                   /*tag= aj
FT                   /note= "KR2"
FT   misc_feature    complement (34606..35749)
FT                   /*tag= ak
FT   misc_feature    complement (28750..28960)
FT                   /*tag= al
FT                   /note= "DH2 (inactive)"
FT   misc_feature    complement (35749..37144)
FT                   /*tag= am
FT                   /note= "AT2"
FT   misc_feature    complement (37145..38296)
FT                   /*tag= an
FT                   /note= "KS2"
FT   misc_feature    complement (38371..38581)
```

```
FT                   /*tag= ao
FT                   /note= "ACP1"
FT   misc_feature    complement (38677..39307)
FT                   /*tag= ap
FT                   /note= "KR1"
FT   misc_feature    complement (39442..40609)
FT                   /*tag= aq
FT                   /note= "DH1"
FT   misc_feature    complement (40609..41842)
FT                   /*tag= ar
FT                   /note= "AT1"
FT   misc_feature    complement (41842..43093)
FT                   /*tag= as
FT                   /note= "KS of extender module 1 (KS1)"
FT   misc_feature    complement (43144..43660)
FT                   /*tag= at
FT                   /note= "ACP of loading domain"
FT   misc_feature    complement (43777..44629)
FT                   /*tag= au
FT                   /note= "ER of loading domain"
FT   misc_feature    complement (44974..46573)
FT                   /*tag= av
FT                   /note= "CoA ligase of laoding domain"
FT   CDS             46754..47788
FT                   /*tag= aw
FT                   /note= "fkbO gene"
FT   CDS             47785..52272
FT                   /*tag= ax
FT                   /note= "fkbP gene"
FT   CDS             52275..71465
FT                   /*tag= ay
FT                   /note= "fkbA gene"
FT   misc_feature    52362..53576
FT                   /*tag= az
FT                   /note= "KS7"
FT   misc_feature    53577..54716
FT                   /*tag= bb
FT                   /note= "AT7"
FT   misc_feature    54717..55871
FT                   /*tag= ba
FT                   /note= "DH7"
FT   misc_feature    56019..56819
FT                   /*tag= bc
FT                   /note= "ER7"
FT   misc_feature    56943..57575
FT                   /*tag= bd
FT                   /note= "KR7"
FT   misc_feature    57710..57920
FT                   /*tag= be
FT                   /note= "ACP7"
FT   misc_feature    57990..59243
FT                   /*tag= bf
FT                   /note= "KS8"
FT   misc_feature    59244..60398
FT                   /*tag= bg
FT                   /note= "AT8"
FT   misc_feature    60399..61412
FT                   /*tag= bh
FT                   /note= "DH8 (inactive)"

  Query Match             36.6%;  Score 37;  DB 21;  Length 77536;
  Best Local Similarity   60.4%;  Pred. No. 0.33;
  Matches   61;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCCGTCCAAGT 60
           | |  || ||||||||     | || |||||| | | |||||| || ||  |   ||||
Db   56473 CCGCACAGATCGCCCGCCACCTGGGCGCCGAGCTCTACGCCACCGCCAGTACCGGCAAGC 56532

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
             |   |   | |  |||   ||  ||||   ||||||||||
Db   56533 AGCACGTCCTGCGCGCCGCCGGGCTGCCCGACACGCACATC 56573
```

```
RESULT 15
AAQ46806
ID   AAQ46806 standard; DNA; 29879 BP.
XX
AC   AAQ46806;
XX
DT   22-DEC-1993  (first entry)
XX
DE   eryA region of S. erythraea chromosome.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement; ss.
XX
OS   Saccarapolyspora erythraea.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..10199
FT                   /*tag= a
FT                   /note= "ORF 1"
FT   CDS             10218..20921
FT                   /*tag= b
FT                   /note= "ORF 2"
FT   CDS             20922..29879
FT                   /*tag= c
FT                   /note= "ORF 3"
XX
PN   WO9313663-A.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US00427.
XX
PR   17-JAN-1992;   92WO-US00427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Donadio S,  Katz L,  McAlpine JB;
XX
DR   WPI; 1993-242804/30.
DR   P-PSDB; AAR44430-32.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin
PT   cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT   microorganisms
XX
PS   Claim 27; Fig 2; 133pp; English.
XX
CC   This sequence represents a fragment of the Saccarapolyspora erythraea
CC   genome, designated eryA.  The polypeptides encoded by this region
CC   are involved in the biosynthesis of the polyketide segment of
CC   erythromycin.  eryA is organised in modules and each module takes care
CC   of one condensation step.  The precise succession of elongation steps
CC   is dictated by the genetic order of the modules.  This fragment may be
CC   specifically altered such that novel polyketide molecules of desired
CC   structure are produced.  Three types of alteration may be produced;
CC   those inactivating a single function in a module which does not arrest
CC   acyl chain growth; those inactivating a single function in a module
CC   which does affect chain growth; and those affecting an entire module.
CC   The mutations may be introduced by gene replacement.
XX
SQ   Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;

  Query Match             35.6%;  Score 36;  DB 14;  Length 29879;
  Best Local Similarity   62.0%;  Pred. No. 0.6;
  Matches   57;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      10 TCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGTGGGCAGCGG 69
           | |||||  |    ||||| |||||||| ||||| ||  |||||  |||||    |   ||
Db   19147 TGGCCCGCCGAGCGGGGGCGGAGGTGTTGGCCACCGCCGGCCCGGCCAAGCACGGGACGC 19206

Qy      70 TTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           | | |||  | ||  |      |  |||||||
```

Db 19207 TGCGGGCGCTCGGTCTCGACGACGAGCACATC 19238

Search completed: November 5, 2002, 14:00:26
Job time : 1436.22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_29800_29900
Perfect score: 101
Sequence: 1 cggtgcaaatcgcccgatgg.....gcgtgccgcgcacgcacatc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	68750	1	AF210843	AF210843 Sorangium
3		101	100.0	68750	6	AR193029	AR193029 Sequence
4		101	100.0	68750	6	AR199551	AR199551 Sequence
5		101	100.0	68750	6	AR199559	AR199559 Sequence
6		101	100.0	68750	6	AR199567	AR199567 Sequence
7		101	100.0	68750	6	AR201097	AR201097 Sequence
8		101	100.0	68750	6	AR208671	AR208671 Sequence
9		101	100.0	71989	6	AR172664	AR172664 Sequence
10		49.8	49.3	41097	1	AF016585	AF016585 Streptomy
11		49.8	49.3	123580	1	AF263912	AF263912 Streptomy
12		49.8	49.3	125401	6	AX211739	AX211739 Sequence
13		48.8	48.3	66808	1	SAU421825	AJ421825 Stigmatel
14		48.4	47.9	30000	6	AX250261	AX250261 Sequence
15		48.4	47.9	30000	6	AX250262	AX250262 Sequence
16		48.2	47.7	41782	1	AY118081	AY118081 Streptomy
17		46.6	46.1	65140	6	AX211705	AX211705 Sequence
18		43.4	43.0	3225	1	AF232752	AF232752 Mycobacte
19		41.8	41.4	104326	1	AB070940	AB070940 Streptomy
20		40.8	40.4	1784	1	AF411573S3	AF411575 Actinomad
21		40.8	40.4	47981	1	AF263245	AF263245 Micromono
22		40.8	40.4	47981	6	AX112026	AX112026 Sequence
23		40.4	40.0	30000	6	AX250263	AX250263 Sequence
24		40.2	39.8	113193	1	AF357202	AF357202 Streptomy
c 25		39.2	38.8	15970	1	AE007061	AE007061 Mycobacte
c 26		39.2	38.8	53450	1	MTV018	AL021899 Mycobacte
27		38.8	38.4	32870	1	AF007101	AF007101 Streptomy
c 28		38.2	37.8	82746	1	AF453501	AF453501 Actinosyn
29		37.2	36.8	18614	1	AE007033	AE007033 Mycobacte
30		37.2	36.8	38000	1	MTCY06H11	Z85982 Mycobacteri
31		37	36.6	37948	1	AF079138	AF079138 Streptomy
32		37	36.6	77534	1	AF235504	AF235504 Streptomy
33		36	35.6	20235	1	SERERYAB	M63677 S.erythraea
34		36	35.6	20235	6	AR049368	AR049368 Sequence
35		36	35.6	20235	6	AR095529	AR095529 Sequence
36		36	35.6	20444	1	SEERYABS	X62569 S.erythraea
37		35.6	35.2	6459	6	AX089458	AX089458 Sequence
c 38		35.6	35.2	16096	1	AE007124	AE007124 Mycobacte
c 39		35.6	35.2	24292	1	MTCY24G1	Z83858 Mycobacteri
40		35.6	35.2	43280	1	SFU78289	U78289 Streptomyce
41		35.6	35.2	50000	6	AX089417	AX089417 Sequence
42		35.6	35.2	50000	6	AX089420	AX089420 Sequence
43		35.6	35.2	80161	1	AY007564	AY007564 Saccharop
44		35.6	35.2	80161	6	AR165018	AR165018 Sequence
45		35.4	35.0	14402	1	AF098795	AF098795 Pseudomon

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDEERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRARERH
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLRPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFCPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRHIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260. .10492
                     /gene="epoB"
     CDS             6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
     gene            10639. .16137
                     /gene="epoC"
     CDS             10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene       16134. .37907
                /gene="epoD"
     CDS        16134. .37907
                /gene="epoD"
                /note="type I polyketide synthase; epothilone PKS modules
                3, 4, 5, and 6"
                /codon_start=1
                /transl__table=11
                /product="epoD"
                /protein_id="AAF62883.1"
                /db_xref="GI:7453560"
                /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29800 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 29859

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   29860 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 29900


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDHPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADPPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKGTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251. .21749
                     /gene="epoB"
     CDS             16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match             100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy     61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
          |||||||||||||||||||||||||||||||||||||||||
```

```
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown'
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  35412 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 35471

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||||||||||||||
Db  35472 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 35512


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-13;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  29800 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 29859

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||||||||||||||||||||||||||||| ||||||||||
Db  29860 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 29900


RESULT 10
AF016585
LOCUS       AF016585               41097 bp    DNA     linear   BCT 07-DEC-1997
DEFINITION  Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)
            gene, partial cds; polyketide synthase modules 1 through 7 (nidA)
            genes, complete cds; and N-methyltransferase homolog gene, partial
            cds.
ACCESSION   AF016585
VERSION     AF016585.1  GI:2558836
KEYWORDS    .
SOURCE      Streptomyces caelestis.
  ORGANISM  Streptomyces caelestis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 41097)
  AUTHORS   Kakavas,S.J., Katz,L. and Stassi,D.
  TITLE     Identification and characterization of the niddamycin polyketide
            synthase genes from Streptomyces caelestis
  JOURNAL   J. Bacteriol. 179 (23), 7515-7522 (1997)
  MEDLINE   98053867
   PUBMED   9393718
REFERENCE   2  (bases 1 to 41097)
  AUTHORS   Kakavas,S. and Stassi,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park
            Road, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .41097
                     /organism="Streptomyces caelestis"
                     /strain="NRRL-2821"
                     /db_xref="taxon:36816"
     gene            complement(<1. .396)
                     /gene="nidi"
     CDS             complement(<1. .396)
                     /gene="nidi"
                     /note="ORF6; similar to tyli"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P-450 hydroxylase homolog"
                     /protein_id="AAC46023.1"
                     /db_xref="GI:2558837"
                     /translation="MVDSVTGPMELSKDANAKELLDWFSHNRTHHPVFWDEGRQAWQV
                     FRYDDYLTVSNHPEFFSSDFTEVAPTPPELEMILGPGTIGALDPPAHGPMRKLVSQAF
                     TPRRMAGQEQRIRVIAEELLDRVRGQKTIA"
     gene            627. .13649
                     /gene="nidA1"
     CDS             627. .13649
                     /gene="nidA1"
                     /note="orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase modules 1 and 2"
                     /protein_id="AAC46024.1"
                     /db_xref="GI:2558838"
                     /translation="MAGHGDATAQKAQDAEKSEDGSDAIAVIGMSCRFPGAPGTAEFW
                     QLLSSGADAVVTAADGRRRGTIDAPADFDAAFFGMSPREAAATDPQQRLVLELGWEAL
                     EDAGIVPESLRGEAASVFVGAMNDDYATLLHRAGAPTDTYTATGLQHSMIANRLSYFL
                     GLRGPSLVVDTGQSSSLVAVALAVESLRGGTSGIALAGGVNLVLAEEGSAAMERVGAL
                     SPDGRCHTFDARANGYVRGEGGAIVVLKPLADALADGDRVYCVVRGVATGNDGGGPGL
                     TVPDRAGQEAVLRAACDQAGVRPADVRFVELHGTGTPAGDPVEAEALGAVYGTGRPAN
                     EPLLVGSVKTNIGHLEGAAGIAGFVKAALCLHERALPASLNFETPNPAIPLERLRLKV
                     QTAHAALQPGTGGGPLLAGVSAFGMGGTNCHVVLEETPGGRQPAETGQADACLFSASP
                     MLLLSARSEQALRAQAARLREHLEDSGADPLDIAYSLATTRTRFEHRAAVPCGDPDRL
                     SSALAALAAGQTPRGVRIGSTDADGRLALLFTGQGAQHPGMGQELYTTDPHFAAALDE
                     VCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARGTQAHLVL
                     GHSVGEITAAHIAGVLDLPDAARLITARAHVMGQLPHGGAMLSVQAAEHDLDQLAHTH
                     GVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKVSHAFHSALMDPMLGAFR
                     DTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQTAHHQGTT
                     TYLEIGPHPTLTTLLHHTLDNPTTIPTLHRERPEPETLTQAIAAVGVRTDGIDWAVLC
                     GASRPRRVELPTYAFQRRTHWAPGLTPNHAPADRPAAEPQRAMAVGPVSREALVRLVG
                     ETTASVLGLDGPDEVALDRPFTSQGLDSMTAVELAGLLGTAAGVALDPTLVYELPTPR
                     AVADHLAKTLLGESAADADQEVNGRTGEAEAKAGDPIAVIGIGCRFPGGVATPDDLWE
                     LVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQFDAEFFGISPR
                     EATAMDPQQRLLLETSWEALERAGVVPESLRGGRTGVFVGTTAPEYGPRLHEGTDGYE
                     GFLLTGTTASVASGRIAYALGTRGPALTVDTACSSSLVALHLAVQSLRRGECDLALAG
                     GTTVMSGPGMFVEFSRQRGLAPDGRCKAFSADADGTAWAEGVGMLLVERLSDAERLGH
```

RVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDAVEAHGTGTAL
GDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGIIKMVQALRHDTLPRT
LHADHPSSKVDWDAGPLQLLTDARPWPADPDRPRRAGISAFGVSGTNAHVVLEEPPAP
EEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIAHALATTRTHF
KHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHRGMGQELYATDPH
FAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARG
TQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHGGAMLSVQAAEHDL
DQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQQLHQQGIRHTWLKVSHAFHSALMD
PMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQT
AHDQRTTTYLEIGAHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTALATLHTTGHT
TTLHTTSPQTHHLDLPTYPFQRDRYWMEPVRVAQVSGQPGADRLRYRVVWEAAPEDHT
VPSAERWLLLGGSDPADAGILVGVEEQLVGHGAAVQRLDIPAPAGRRLLPSCSELPVR
GPSHTGEFCGSRLRSPRVAEAVALLQALGDAGADAPLWIATRGAVAAQAGEAPSVGGA
QLWGLGQVAGLELADRWGGLVDLPADPAPAALRGLARVLVSNAPDNQVAIRASGVEVR
RVVPAPGRPVRNDWAPSGTVLITGGTGALGSQVARRLALAGAPHLLLAGRRGNSGAAE
LVDELTALGAEVTVAACDAADRDALVSLLATIPEHRPLTAVLHAAGVLDDGVLDGLTP
ERIDAVLRAKATAARHLDELTADLDLDAFVLFSSIVGVWGNGGQATYAAANAALDALA
HGRRARGQRATSIAWGPWAGSGMAAGDGAKSLARDGIDVLEPEEALDLLEQAVRSDET
TLLVADVDWETFLSSSASRRFQLPLFGRIPAARAVAVSGADGSPDDVPVWLSPQLSAA
ERQLRLLDLVRTEAASVLRHGSATAVDAETAFRAGGFDSLTLVELRNRLASATGLRLS
NTLLFDHPTPAAVAEYLGEQLLGTEPSAAASHGDAGARTVVASDDPVAVVGMACRYPG
GVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQF
DAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGASQQEYGT
QSREAADKYGGHLLTGTLASVMSGRVAYTLGLQGPALTVDTACSSSLVALHLAVQSLR
RGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVGVLLVE
RLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDA
VEAHGTGTPLGDPIEAEALLATYGHPERQTPVWLGSLKSNIGHSQAAAGIAGIIKMIQ
AMHHGTLPRTLHVDAPTSQVDWEAGSLQLLAEARPWPADPDRPRRAGISAFGVSGTNA
HVVLEEPPAPEEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIA
HALATTCTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHPGM
GQELYTTDPHFAAALDEICEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQT
ALYRTLTAHGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHDGAM
LSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKV
SHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARH
TVRFADAVQTAHDQRTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTA
LATLHTTGHTTTPHPSHIPAQRVSLPAYPFQRRAYWMPNSAAHIGRSDAEAATRLGLA
RQDHPFLTGATSIVGSGATLLTGRVSLAAHPWLADHTVAGAVLFPGTAFADLLLHVAE
ETGSGTVEELTLHVPLVLPAQGAVQLQVLVEAPDERGRRAATVAARPEEETGAEWTTH
AAGVLAGAEPAAPAVDWADWAAGAWPPPGAEAVDVDELYAEFAAAGYDYGPAFTGLTG
AWRRGEEVFADVRLPEAVAGSGEEERRFGVHPALFDAALHPWRVGGPWQESGPDADGG
QSTLLPFSWQGLTLYGTGAEQVRVRLAPSDGGAFSVRAADPAGHPVLALDALVLRRVP
SSTPGALPAAPMYHIDWQPVDSTSLVDTRRYAVVGADADETAEVLRAFLPDAHTTAHA
DLAALRAAPDAVAPTPGPALVVALIPGSPPPAPEATSAPDPVRATLAHGLSLVQALIE
DERLTTGTTLAVVTRGAVEATQTDVPDLAGAALWGLVRSAQSEYPGRFRLIDVDDSPE
SRAALATALASDEPQLALRAGVMLAPALTPAEAAGSPAPADAAPTAHGMDGTVLLTGC
TGALGRRVAVHLARRHGVRRMLLVSRSGPNAPESAALERELTELGVHTSFVSCDLADP
ADVRKAVAGVPSDHPLTGVVHTAGVLDDGALARLTPERIDTVLRAKADAVRHLHEATL
GHPLRAFVLFSAAAGLLGRPGQASYAAANAVLDAFARLRRAEGLPAVSLAWGLWDERE
GMAGGLDATALRRLRRDGILPMPADQALELLDQALTTYRDGPALLVPLLLDSTALRRT
AGENGPASVPPLLRSLVPSHPRRHRTTTATSVRDESGPGPSARLAALPAKERMTVLLD
TVSEQVAGVLGHASAAEIDPERPFHEIGFDSLATLELRNRLGRLVGLRLPSTLAFDHP
TPRKVAEWINAELPAANPALGTSAEDPVLTGIDQMARAIALMDAEDGRRADVRQRLVG
LLTALDRPASGADSTRAAATTVADRLDEATDDEIFAFLDEQLEEQPDEPLGAPLDERL
DEQP"

gene 13693. .19212
/gene="nidA2"

CDS 13693. .19212
/gene="nidA2"
/note="orf2"
/codon_start=1
/transl_table=11
/product="polyketide synthase module 3"
/protein_id="AAC46025.1"
/db_xref="GI:2558839"
/translation="MTAENDKIRSYLKRATAELHRTKERLAELESARHEPVAVVGMAC
RYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHD
AAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPNTLRTTRTGVFTGVMYD
DYGSRCFPSLPEVEGYVVNGSAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLAV
QSLRRGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKSFADTADGTAWAEGVGV
LLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSAD
DIDAVEAHGTGTPLGDPIEAGALLATYGHPERQTPVWLGSLKSNIGHTQAAAGIAGII
KMIQAMHHGTLPRTLHVDHPSSKVDWKAGSLQLLAQARPWPADPDRPRRAAVSAFGVS
GTNAHAIIEEPSTLIEAPVPDETPVPAHTPPVIAWPLSAHTPTALRAQAARLDSWLHH
TQADPQDIAHALATTRTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFT
GQGAQHPGMGQELYTTDPHFAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYT
QPALFALQTALYRTLTARGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLM
GQLPHGGAMLSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQ
NVRHTWLKVSHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTP
DYWIDHARHTVRFADAVQTAHHQGTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREH
PEPETLTTALATLHTTGHTTTLHTTSPQSHHLDLPTYPFQRDRYWMAVPPRAAVGDLA
AAGVTAVGHPLLAAVTEVPETDTTVWTGMLSARALPWLADHLVWDRGVAPGTAVLEIV
LEIGERIGLPCIGELVLEAPLSWGADAPAVVRTVVGALSPEGTREVSLYSRPEDEAVW
TRHASGRLARAGSEEMGSGVFAELSGVWPPPGAQEVDISGQYELFAAAGVGYGGAFRG
LRAVWRREDEIFAEVRLPDDHATDAAGYGLHPALLDAALHPIALLDPLGDGGHGLLPF
VWTDVLRPGNGGHTLRVRVTRAQDPADGGQDDRPGGAVSLAAADRDGVPVFVARSLVL
RRVSATQLPAAPFVPLHRLDWVPIPGFTESTGAGGSDPSVLDLTELTADVHLTVREVL
EAVQAHLREPVGDGGRLVVLTRGAAGPEGGADAPAGAAAWGLVRSAQSEHPGRFVLID
VDGTEESSRALPAVMAAVGDVDAPQVAVRQGRPSVPRLVPLTADRTEAEPGGPGSAGL
LGPDGTVLITGGTGALAAVTARHLVARHGARHLLLVSRRGTDAPQAAELASELGALGA
DVEFRACDVTDRAAVQRLLAGIPAEHPLSCVVHTAGVLDDGVISAQTPERVAAVLGPK
ADGALHLDELTRDRERSVPLVLFSSVSGVLGSTGHAGYAAANAFLDALAARRRADGHP
ALSLGWGWWTGTGMAAGLDRADTARIRRSGIAPLDAETAMDLFDRALARPEPALLPVR
LDHAALREAARAGTLPEVLRDLAGVPSRAHPAPGSEGTATSTGTGSGLAERLARRTVA
ERLPLLLELVRTEVAAVLGHGDPGAVGAERSFKDAGFDSLTAVDLRNRLNARTGLRLP
ATLVFDHPTPLALAELLLAGLDVEPAADGPEDLAAAFDRLARGLAALDDRAARVGAAE
RLRSLLTALTSADPATGTGAQAGPEADGALTEVVTDRLRSASDDELFDLYDSEFR"

gene 19269. .30233
/gene="nidA3"

CDS 19269. .30233
/gene="nidA3"
/note="orf3"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 4 and 5"
/protein_id="AAC46026.1"
/db_xref="GI:2558840"
/translation="MSATNEEKLREYLRRAMGDLHSARERLAELESARHEPVAVVGMA
CRYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLH
DAAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGVMH
HDYGSHQVGSAADASGQLGLGTAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLA
VQSLRRGECDLALAGGTTVMATPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVG
MLLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALADAHLTP
DDIDAVEAHGTGTPLGDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGI
IKMVQALRHDTLPRTLHADHPSSKVDWEAGPLQLLAQARPWPADPDRPRRAGISAFGV
SGTNAHAIIEEPPTLTEAPADTHTPPVIAWPLSAHTPTALRAQAARLRARLHRTSDAL
SLADSAGIGHGLAVGRAALPHRAVLLGDGAAPLDALAALASGEVSPDVVTGSAADVRR
VAFVFPGQGAQWAGMGAELLDSSPVFAAELARCEAALEPFVDWSLTDVLRGAPGAPGL
DRVDVVQPVTFAVVVALAAMWRWLGVEPAAVVGHSQGEIAAAHVAGVLSLEDAARVVA
LRSQLIARELAGRGSMASVALAAADVESRLAGAEAGGGVRDVEIAAVNGPETTVVCGA
PGAVDSLLGVLQGEGVRVRRIDVDYASHSRHVEGIRDELAAVLAGLRPRAGRVPFYST
VEAEPLDGTALDAGYWYRNLRQRVRFESALRAMLADGVDAFVECSPHPVLTVPVRQTL
EDAGAGAVAVGSLRRDDGGLRRFLTSAAEAQVAGVPVDWAALCPRAGWVDLPTYAFQR
ERYWVAPAEPGPAAGAGSAAATGPAAATAARSGSGQEDAASAYGDRLAYHVAWQGLLG
GADAWRPGPRLLIVPAGEREAEVADSVEQAIASFGGTVRRVTADPLSTGRDGLRALLA
PVVAGEEPLTGAVSLLGLCTDGHPEHPGVPAGVTGTLALVQALADLGVAVPLWTVTRG
TVLVVGDSTTVPARLLRSLLDDGAERVVLAGADAGTATDAGTAELAPDAAVVRIPCDV

Query Match 49.3%; Score 49.8; DB 1; Length 41097;
Best Local Similarity 68.3%; Pred. No. 0.05;
Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           | ||||| |||||||| |  || ||  ||||| | || || |||||||| ||||||
Db   28327 CCGTGCAGATCGCCCGCCATCTGGGCGCGCAGGTGCTGGCGACCGCGAGCCCCACCAAGT 28386

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
             ||| ||  | |  | | ||||||| ||  |  |||||||
Db   28387 GGGACGCCCTGCGCGGCCTGGGCGTCGCGGACGAGCACATC 28427
```

RESULT 11
AF263912
LOCUS AF263912 123580 bp DNA linear BCT 24-MAY-2000
DEFINITION Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster, complete sequence.
ACCESSION AF263912

```
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELPLPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNPAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPLDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDHRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
gene                 4714. .5748
                     /gene="nysDIII"
CDS                  4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
gene                 5930. .34363
                     /gene="nysI"
CDS                  5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
```

```
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                     QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                     LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                     DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                     SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                     PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                     AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                     SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                     LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                     SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                     ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                     VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                     PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                     VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                     TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                     PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                     CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                     LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                     VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                     TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                     ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                     AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                     DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                     ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                     LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                     AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                     QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                     AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                     RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                     ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                     PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                     ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                     TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                     DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                     VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                     NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                     GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                     ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                     AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                     AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                     EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                     SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                     LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                     EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                     FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                     TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                     IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                     PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                     AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA

  Query Match             49.3%;  Score 49.8;  DB 1;  Length 123580;
  Best Local Similarity   68.3%;  Pred. No. 0.043;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           || | || ||||||||     | || ||||||||||||||||||||| ||   |  |||||
Db   39095 CGATCCAGATCGCCCGGCATGTCGGTGCCGAGGTGTTCGCCACGGCCAGTGAGGGCAAGT 39154

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  | | | |   || |||||||| |   |   ||||||
Db   39155 GGGACGTGCTGCGCTCCCTGGGCGTGGCCGACGACCACATC 39195


RESULT 12
AX211739
LOCUS       AX211739              125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN

  Query Match             49.3%;  Score 49.8;  DB 6;  Length 125401;
  Best Local Similarity   68.3%;  Pred. No. 0.043;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           || | || ||||||||     | || ||||||||||||||||||||| ||   |  |||||
Db   39503 CGATCCAGATCGCCCGGCATGTCGGTGCCGAGGTGTTCGCCACGGCCAGTGAGGGCAAGT 39562

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCACATC 101
           |||  | | | |   || |||||||| |   |   ||||||
Db   39563 GGGACGTGCTGCGCTCCCTGGGCGTGGCCGACGACCACATC 39603


RESULT 13
SAU421825
LOCUS       SAU421825              66808 bp    DNA     linear   BCT 16-MAR-2002
DEFINITION  Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
            ORF1, stiA gene, stiB gene, stiC gene, stiD gene, stiE gene, stiF
            gene, stiG gene, stiH gene, stiJ gene, stiK gene, stiL gene and
            ORF9.
ACCESSION   AJ421825
VERSION     AJ421825.1  GI:19572309
KEYWORDS    acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
            kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
            ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
            channel beta chain; ribosome binding factor; stiA gene; StiA
            protein; stiB gene; StiB protein; stiC gene; StiC protein; stiD
            gene; StiD protein; stiE gene; StiE protein; stiF gene; StiF
            protein; stiG gene; StiG protein; stiH gene; StiH protein; stiJ
            gene; StiJ protein; stiK gene; stiL gene.
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1
  AUTHORS   Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
            Hofle,G. and Muller,R.
  TITLE     The biosynthesis of the aromatic myxobacterial electron transport
            inhibitor stigmatellin is directed by a novel type of modular
            polyketide synthase
  JOURNAL   Online Publication
  REMARK    J. Biol. Chem., 10.1074/jbc.M111738200
REFERENCE   2  (bases 1 to 66808)
  AUTHORS   Muller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
            Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .66808
                     /organism="Stigmatella aurantiaca"
```

```
                     /strain="Sg a15"
                     /db_xref="taxon:41"
     CDS             complement(4978. .6231)
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450 monooxygenase"
                     /protein_id="CAD19077.1"
                     /db_xref="GI:19572310"
                     /translation="MLRSVSASRPPSPANPVAAFDLARLDDAFYADPFPLYRAMRERD
                     PVHRMPDGSLFLTRWADLDCIYRDTRTFSSDKRAEFGAKYGDAPLFEHHTTSLVFNDP
                     PLHTRVRRLIVGALTPRALSTMEPGLRTLVDRLLDGLAVKGAADLIEDFAAAIPIEVI
                     GNLLDIPTEERGPLRGWSLAILGALEPRLTAEQEACGNEAVTEFLDYLRILVAQRRAR
                     PGDPATDVLTRLIQGESDGERLTETELLHQCVFLLNAGHETTTNLIGNALELLARFPD
                     ERARLLRAPALIPTAVEEVLRYESSNQLGNRRVAEDTEIGGVAVPTGTFLTLCIGAAN
                     RDPARFEDPEHFDVGRQDNRHLAFAGGAHTCAGMNLARIEARIALAAFLARFPDYALT
                     APPVRARRARFRGFTAMPTRLGSLR"
     CDS             6369. .7730
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAD19078.1"
                     /db_xref="GI:19572311"
                     /translation="MEELISEALPDGSRVGPWRIEGRAGYGTYGAVYRAHRVGRICPR
                     SVALKLARYPNDLRFERHAELLERVRHANVPRLLGHGTWRAGPQGEAHPYVVMQWVEG
                     VRLYQWNKERLRTAREVLRVLAQVARALEATHMRGGLHRDVKGENVLVSPEGHASLMD
                     FGCGTWEGAAPLTGGILAPGTKAYRSPEALRFHWNHRNEADAHYQATPADDVYALGVT
                     AYRLCTGIYPPPATEPSGVGEDERETQEVKKPPSQVRPLVPALEQLILRMLRDNPKER
                     GRAGELAAEMEALAERAGAEADEPMNPNHSSVPSEITGRPVSVQREWGGWPVALVLVA
                     VGLLLWGSGSLNTGRRVAPASRMTEGGTGGVADAAVEALPVSSAAREPEPSGLNLDML
                     KEPLPGQRRPPCPRFQISIQGGCWLEVPQSSPPCGESYYEWKGACYSPVAAPPRPRTS
                     DPQ"
     CDS             complement(7849. .8577)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl transferase"
                     /protein_id="CAD19079.1"
                     /db_xref="GI:19572312"
                     /translation="MSQNIYDDPSFFESYSQLARSLKGLAGAPEWPALRALLPDVKDL
                     RVVDLGCGFGWFCRWARKNDAAHVLGLDLSEKMLTRAKAETSGDAITYQQADLERLVL
                     PEGAFDLAFSSLAIHYVEDLEPLLATVHRGLTPGGWFVFSIEHPIFMASRQPEWMTGP
                     EGHRTWPVDSYQLEGKRTDHWLAEEVVKYHRTLGTTLNALLEQGFTLRHIEEWGPTDA
                     QLSAHPELAEERDRPMFLLVSAQR"
     CDS             complement(8622. .8879)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA binding protein"
                     /protein_id="CAD19080.1"
                     /db_xref="GI:19572313"
                     /translation="MALEDFRSAQERVKALKTRPSNDRLLELYSLFKQATEGDVQGK
                     RPGILDVKERAKYDAWVGHKGLGRESAMQQYVALVERLLRG"
     CDS             8961. .10070
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="potassium channel beta chain"
                     /protein_id="CAD19081.1"
                     /db_xref="GI:19572314"
                     /translation="MVYLLDRSRSQLRWILNGSAHRVRFAWGGHARPRRMHFRHLGRS
                     GLMVSEISYGNWLTHGSQVEAAAVACVRAALDEGITTFDTADVYAGTRAEEVLGRAL
                     QGQRREGLEIFTKVYWPTGPGPNDRGLSRKHILESIHGSLRRLKTEYVDLYQAHRYDT
                     ETPLEETMQAFSDVVRQGKAHYIGVSEWTADQLRESVKLSRQMGFQLVSSQPQYSMLW
                     RVIEAEVVPASQELGVGQVVYSPIAQGVLTGKYRPGQKPPEGSRATDEKGGASFIKRF
                     LNDDVLTRVQQLTPVAQEAGLSLAQLAVAWVLQNPNVSSAIIGASRPEQVKENVKAAG
                     VKLSEEAMRHIDTLLGPIVERNPKKTLENSPRKRP"
     CDS             10428. .10862
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosome binding factor"
                     /protein_id="CAD19082.1"
                     /db_xref="GI:19572315"
                     /translation="MTSFKQRRRRAPRRGTESSLSSFEAQSLPAHHLRVQSTLFQEVA
                     RLFRGELSDPLLEGVSVTSFELTPDGRNARIGFTLPPEYAEGGSAPIEAALVRASGFI
                     RAQLALGLNLKRVPHLRFVCVGVAPPGFSIPEDGSAEEGGES"
     CDS             10859. .12016
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD19083.1"
                     /db_xref="GI:19572316"
                     /translation="MRPRRIEGAPGAVPILAWARAVPAGAVKQLQLLARQPYLAEHVA
                     AMPDVHVAEGVAVGTVFATERHIVPGALGGDLGCGVSAVRFDFPAMSLGRVELTRLLA
                     AFARAVPVGDAVHRGRGLSLPPGLDIGRLSTHRLQREWERLVPKHLGTLGGGNHFLEL
                     DRDAEGALWLLLHTGSRGVGAAIAAHHLRVAQVSGEGSLPGLRTDTSEGAACLADTEL
                     ACQFARANRDTLLQRAVEVLAGAIGQGPVLDSRVDVHHNHVAAESHFGRTLLVHRKGA
                     VGLVAGQQGLIPGSMGTASYLVAGKGEPQAFCSCSHGAGRVMTRREARERIRPAALAH
                     ALRRVVFDEGRVTALVEEAPTAYRDITEVLEDEAELVTPLRRLTPIAVLKG"
     CDS             complement(12082. .13974)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="cellulase"
                     /protein_id="CAD19084.1"
                     /db_xref="GI:19572317"
                     /translation="MKNSLLLLMLAVLPLYGTAALARPGMANAEPGHFEGPENAGLTA
                     AQCNPRVPLSTRGRYIVDRCGERFKLKSVNWFGASDQLEVVGGLDKQKLSDLVAGMKA
                     MGFNSVRLPFSNNMLHPDDNKPAAQRCLEKYSVTCIDSTKNPELLNKTALEVYDAVVA
                     ELTRQGLVVILNNHTTKSMWCCGWDGNGFWDSSQALQRWQDDWVMLAARYQGNKWVAG
                     ADLRNEVRPDGLDSPNWGMRNQHDWHMAAQTMGNLLLRTNPDLLIVVEAVNWWGLLDG
                     SRPQLKPVRQRPVALLRGDKLVYAVHNYGYTGPNQSGGSLGSGPKYSDMDKPTLYGTL
                     DQEWGFALAANQAYTAPVWMSEFGVGYNEQAANSRAWFSNLADYLIEKDVDWAYWAIN
                     AAKLQNSVQGEDETYGLWSYPDWNGVRSGDWRLGTGPLGRLLGADGRTGAVDATRFLP
                     MTVLLKDGSSTYYVDNNSLDSDWHSGKAKISCPRGYRMVGVSANFTILCTNAGAVPAQ
                     QGTGSYQTITQEVSPLRYPGDWANQSIKFECPTGFYAVGISTANPFFLELAGLLCEQN
                     TGGIPLNSKSAKNFWGGDQRASSSGGDWSVSQYKGQCADNQYLIGLAHRRSGLADYPS
                     VVLCSN"
     gene            14345. .21466
                     /gene="stiA"
     CDS             14345. .21466
                     /gene="stiA"
                     /function="PKS"
                     /codon_start=1
                     /transl_table=11
                     /product="StiA protein"
                     /protein_id="CAD19085.1"
                     /db_xref="GI:19572318"
                     /translation="MTQLASRLGVGSESIDPRERFSRYGIDSRDTKSLLAELGALLKR
                     PLSPTLAWTYPTVEALSRHLAGEADVSGGERADTLGVDQDEPIAIIGMAGRFPKAPDV
                     ESFWSLLRDGVDAITEVPPERWDLDAFYSVGLKSPGKMNTRWGGFLDQVDQFDPLFFG
                     ISPREAIQMDPQQRLALELSWEALESAGVAVDSLMDSRTGVFMGAMWSDYAKLTRGAL
                     GLIGPHTATGQDSSIIAARVSYFLGLRGPSLTVNTACSSALVAVHLACQSLRNGEAAM
                     ALAGGVNLLLSPESSVAMSQLGALSPDGRCKAFDARANGYVRGEGGGVVVLTPLSRAI
                     ANGYPVLCVIRGSAVNNDGFSNGLTAPNPRAQEEVLRDAYARAGVSPDKVHYVETHGT
                     GTVLGDPIEANALSAVLGVNRPAERPLILGSVKTNIGHLEAASGIAGLMKVVLSIQHR
                     AIPPNLHFEQPNPYIRFEELRLLVPKSLHPWPNEGESALAGISSFGFGGTNCHLVLEA
                     PSPGSAHIAPLAAETPEELSLLARQARDAVSVVGSQVELADLCFSVAARCREGSQRLA
                     LIARSPAELAAQVDGFLAGEKRPGLLRGEVTSGGSQKPVFVFSGTGSQWPGMGRELLR
                     TEPVFRGAIERCDRALRGLSGWSMLEELSAGEASSRLKECSGLQTAIFAIQVGLAALW
                     RSWGIEPAGVVGHSIGEVAAAHVAGILNLEDAARVIHHLSRLMGRLSGQGAMIVVDLS
                     AEEANRLLREWERHVVVAACNAPASTVLSGAREVIEELGAELERRGIAAHTVRIDTPS
                     HSPLMDPFMGELAEVLAGLSPRVAAVPMVSTVTGAFLKAEECTAAYWVRNLREPVLFS
                     QALTKLFEQDLRLFLELSPHPILARSVEQCLKGEGRQGVVLASMRRGENEGACMRESL
                     GRLFTLGTSVAWQALYPAEAKRVELPMGLGTGSLDAELRGSARLLPLSAHTAEALEDA
                     ARALQASIAKHPDVRLDDLCYSASVGKSVHAHRLAFVTDSKRALVEHLDDFLKGESRP
                     GRIAGVCATGEPKKVVFVFPGQGSQWQGMGRQLLRQEPVFREAIEECDAAIRSCAGWS
                     LIDALQADEPHSRLGSIEVIQPTLFAIEVALAALWRSWGVVPDAVVGHSMGETAAACV
                     AGALRLTDAARVICERSQLLRRVAGQGGMALVELSLEQARAEMLSRKDGRVSIGASNSP
                     RSTVLSGDAIALEEILRELEAQGTFCRRVKVDVASHSPQVDPLKDDLLRVLSTIAPRA
                     ATVPLYSTVTGEPIEGARMDASYWMRNLREPVLFGQAIERLAGDGHGVFVELSPHPIL
                     LPSIDQTLRHKSHKGVVLPSLRREEPERAVLLESLGTLFTGGGRVDFRGLHPLGGHRV
                     RLPHLPWQRQRFWLDMDVAAAGAASVGAAWHPGGGSEHPLLGHRLPSLAHAPACHVWE
                     NKLSRSAVSYLDDHRVGGTVVLPGTSYIEMALAAAAKGSGKKTVSLRDVELREKMSFL
```

```
                     ETASRTVQLSLIQEEAGHSSFRLHGRSTPDEPWTLHGGGTVHVEEGDGPPSADPVLWA
                     AVETVRLQGKEVSASDAYEAMHAQGNEYGPLFRGIEQMWVRDAELVARLRVPEPLEQE
                     LPEYRFHPAALDLGLQALIVLAAEHGCEGLALPTRLESIRAHASPGARMWCHVRLNPV
                     REAASFSGEVLFLDESGQRLAELQGARFQILGEGSRPHDGGRLVDWLHEVEWQPAPVP
                     VPTAEPSSQGRWVLFADGTGVAEGLAAQIAARGEPCIRVLPGMAFAEAEDGHFVVDPE
                     QPEDVRKLFGALSRTSTIPLRGIVHLWSLDAPSPEEGGVDALSRARRLGSGSVLSLIQ
                     GLNAAELAQKPRLWLVTRGVQAVGSAGPGAGAVAQSPLWGLGQVVAYEHAELRCTRVD
                     LAPVSVPGEEQTLLQELLAPAGDDQVALRGRERHVPRLTRCRTDEVAGRSTDLRAGAI
                     RSDGCYLITGGLGGLGLTIAKWLVARGARHLVLVGRRGASDAAEIALGQMRGAGAECL
                     VMKADVANTSEVEELVGTIQRKMPPLRGIFHAAGLLDDGTVLQLTQARLRSVMAPKVE
                     GAWNLHRATLGASLDFFVLFSSAASLVGIPGQGNYAAGNAFLDALAGLRKARGLACQS
                     IQWGPWSEIGLAAEQDNRGERLALRGMGSLNPREGLEALEALLSPGEPLRERARIGVL
                     RLDARRWCKSYPAIRSSPLFSVLLAGERNAKQVLERSGLRARLNEEPARTRRAALESF
                     LREEAAQVLRLPPARIDLSRSLQTLGMDSLLSLELRNRLEANLEITLPATVLWNYPTL
                     SALAPFLAERMGIQLNGVEEPPQAVESPSVSEEDGAAMLKMLEQFQELSNEDLLGMLS
                     DDKEDEVSS"
     gene            21463. .26220
                     /gene="stiB"
     CDS             21463. .26220

  Query Match           48.3%;  Score 48.8;  DB 1;  Length 66808;
  Best Local Similarity   70.7%;  Pred. No. 0.079;
  Matches   65;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| |||||    |    || ||||||||||||||| |||||  ||   ||||
Db   48416 CGGTGCAGCTCGCCAAGCGCGCGGGTGCCGAGGTGTTCGCCACCGCGAGTTCGGGGAAGT 48475

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGC 92
           |||  | | || |||||||||| ||   || |
Db   48476 GGGATGTGCTTGAGGCCATGGGTGTCGAGCAC 48507


RESULT 14
AX250261
LOCUS       AX250261               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 1 from Patent WO0168867.
ACCESSION   AX250261
VERSION     AX250261.1  GI:15984063
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 1 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 1 to 30000 of the monensin biosynthetic
                     gene cluster"
BASE COUNT     3944 a  10629 c  11136 g   4291 t
ORIGIN

  Query Match           47.9%;  Score 48.4;  DB 6;  Length 30000;
  Best Local Similarity   68.4%;  Pred. No. 0.11;
  Matches   67;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || ||||||||||||||||| |||||  |||  ||||
Db   27973 CGGTGCAGATCGCCCGGCATGTGGGTGCCGAGGTGTTCGCCACCGCGAGTCCGGGCAAGC 28032

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCAC 98
           || | | |  |||  |||||| |      | | |||
Db   28033 ACGCCGTGCTGGAGGAGATGGGCATCGACGCCGCCCAC 28070


RESULT 15
AX250262
LOCUS       AX250262               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 2 from Patent WO0168867.
ACCESSION   AX250262
VERSION     AX250262.1  GI:15984064
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 2 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 30001 - 60000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     3911 a  10993 c  10862 g   4234 t
ORIGIN

  Query Match           47.9%;  Score 48.4;  DB 6;  Length 30000;
  Best Local Similarity   68.4%;  Pred. No. 0.11;
  Matches   67;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 CGGTGCAAATCGCCCGATGGATAGGGGCCGAGGTGTTCGCCACGGCGAGCCCGTCCAAGT 60
           ||||||| ||||||||     | || ||||||||||||||||| |||||  |||  ||||
Db   22592 CGGTGCAGATCGCCCGGCATGTGGGTGCCGAGGTGTTCGCCACCGCGAGTCCGGGCAAGC 22651

Qy      61 GGGCAGCGGTTCAGGCCATGGGCGTGCCGCGCACGCAC 98
           || | | |  |||  |||||| |      | | |||
Db   22652 ACGCCGTGCTGGAGGAGATGGGCATCGACGCCGCCCAC 22689


Search completed: November  5, 2002, 12:36:59
Job time : 650.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300
Perfect score: 101
Sequence: 1 gagaccgaacgccgatggac.....gaggtgctggaggaccgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		30.4	30.1	259	10	US-09-923-876-1012	Sequence 1012, Ap
2	c	29.6	29.3	1500	10	US-09-789-836-8	Sequence 8, Appli
3	c	28.8	28.5	239	10	US-09-923-876-2938	Sequence 2938, Ap
4	c	28.6	28.3	656	10	US-09-925-301-122	Sequence 122, App
5	c	28.4	28.1	531	10	US-09-864-761-12628	Sequence 12628, A
6		28.4	28.1	1701	10	US-09-933-814-1	Sequence 1, Appli
7		28.4	28.1	1701	10	US-09-824-134-1	Sequence 1, Appli
8		28	27.7	789	10	US-09-808-387-29	Sequence 29, Appl
9	c	28	27.7	1408	10	US-09-917-800A-1719	Sequence 1719, Ap
10		28	27.7	1446	10	US-09-815-242-7925	Sequence 7925, Ap
11	c	28	27.7	1739	10	US-09-789-836-10	Sequence 10, Appl
12	c	28	27.7	2061	10	US-09-925-299-166	Sequence 166, App
13	c	27.8	27.5	224	10	US-09-923-876-1480	Sequence 1480, Ap
14		27.8	27.5	464	10	US-09-764-887-416	Sequence 416, App
15		27.8	27.5	464	10	US-09-764-887-457	Sequence 457, App
16	c	27.8	27.5	528	10	US-09-793-232-1	Sequence 1, Appli
17	c	27.6	27.3	975	10	US-09-815-242-4062	Sequence 4062, Ap
18	c	27.6	27.3	3124	10	US-09-925-302-340	Sequence 340, App
19		27.4	27.1	1486	10	US-09-122-383-1	Sequence 1, Appli
20	c	27.4	27.1	1841	10	US-09-822-849A-11	Sequence 11, Appl
21		27.4	27.1	13842	10	US-09-861-289-30	Sequence 30, Appl
22		27.4	27.1	36778	10	US-09-861-289-5	Sequence 5, Appli
23	c	27.2	26.9	849	10	US-09-815-242-7917	Sequence 7917, Ap
24		27.2	26.9	2346	10	US-09-860-868-3	Sequence 3, Appli
25		27.2	26.9	2379	10	US-09-764-870-27	Sequence 27, Appl
26		27.2	26.9	3521	10	US-09-860-868-1	Sequence 1, Appli
27		27	26.7	375	10	US-09-960-352-13368	Sequence 13368, A
28	c	27	26.7	687	10	US-09-815-242-7934	Sequence 7934, Ap
29		27	26.7	933	10	US-09-815-242-7970	Sequence 7970, Ap
30		27	26.7	35100	10	US-09-782-378A-26	Sequence 26, Appl
31		26.8	26.5	4039	10	US-09-205-448-7	Sequence 7, Appli
32		26.8	26.5	15872	10	US-09-861-289-1	Sequence 1, Appli
33	c	26.6	26.3	1815	10	US-09-758-269-13	Sequence 13, Appl
34	c	26.6	26.3	11220	10	US-09-861-289-32	Sequence 32, Appl
35	c	26.6	26.3	36778	10	US-09-861-289-5	Sequence 5, Appli
36	c	26.4	26.1	194	10	US-09-864-761-24339	Sequence 24339, A
37	c	26.4	26.1	249	10	US-09-923-876-6100	Sequence 6100, Ap
38	c	26.4	26.1	295	10	US-09-294-093B-3038	Sequence 3038, Ap
39		26.4	26.1	489	10	US-09-822-485-31	Sequence 31, Appl
40	c	26.4	26.1	512	10	US-09-864-761-7634	Sequence 7634, Ap
41		26.4	26.1	817	10	US-09-833-381-1377	Sequence 1377, Ap
42		26.4	26.1	995	10	US-09-822-485-34	Sequence 34, Appl
43		26.4	26.1	1027	10	US-09-803-589-13	Sequence 13, Appl
44		26.2	25.9	575	10	US-09-728-952-21	Sequence 21, Appl
45		26.2	25.9	14759	10	US-09-952-360-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-923-876-1012
; Sequence 1012, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 1012
; LENGTH: 259
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700157970H1
US-09-923-876-1012

Query Match 30.1%; Score 30.4; DB 10; Length 259;
Best Local Similarity 61.2%; Pred. No. 3.5;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy      12 CCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTT 71
           | | ||| ||  | |   || ||| | ||||  |||| | ||||||||||| |  || |
Db       8 CGGGTGGGCGGACAAGATCCACGGCCTGGTGGCGCCGGCCGACGGCGCGCACCACGTGCA 67

Qy      72 GCCGCGCCTCGAGGTGCTGG 91
           |  ||  | ||||  | | |
Db      68 GGTGCTGCACGAGCCGGTCG 87
```

RESULT 2
US-09-789-836-8/c

```
; Sequence 8, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
;  APPLICANT: BRIGHAM, KENNETH L.
;  APPLICANT:  STECENKO, ARLENE A.
;  APPLICANT:  SEALY, LINDA
;  TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
;  FILE REFERENCE: N-6977
;  CURRENT APPLICATION NUMBER: US/09/789,836
;  CURRENT FILING DATE:  2001-02-20
;  PRIOR APPLICATION NUMBER: 60/183,584
;  PRIOR FILING DATE: 2000-02-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 1500
;   TYPE: DNA
;   ORGANISM: Mus sp.
US-09-789-836-8

  Query Match             29.3%;  Score 29.6;  DB 10;  Length 1500;
  Best Local Similarity   61.8%;  Pred  No. 4.9;
  Matches   47;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      16 TGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCG 75
           | | ||  |   ||| |||| |  | | ||| | | |  |||||| ||   |||| | ||
Db     739 TTGGCGGGCGCGGCCGGCGGCCCCGCGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCG 680

Qy      76 CGCCTCGAGGTGCTGG 91
             | ||||  |||| ||
Db     679 GGGCTCGGCGTGCCGG 664


RESULT 3
US-09-923-876-2938/c
; Sequence 2938, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 2938
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700161420H1
US-09-923-876-2938

  Query Match             28.5%;  Score 28.8;  DB 10;  Length 239;
  Best Local Similarity   60.0%;  Pred. No. 9.1;
  Matches   48;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      22 ACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTC 81
           | || || || |  || |||    | |   ||||| ||    | ||    ||| |||| |
Db     219 AGCCGCGTCCTCCATCAGGTTGTACAGCGGGACGGAGCCGGAGCCGGCCAGCCACGCCGC 160

Qy      82 GAGGTGCTGGAGGACCGCGC 101
           ||||| ||| | | || |||
Db     159 GAGGTACTGCACGCCCACGC 140


RESULT 4
US-09-925-301-122/c
; Sequence 122, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 122
;   LENGTH: 656
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (41)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (218)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-122

  Query Match             28.3%;  Score 28.6;  DB 10;  Length 656;
  Best Local Similarity   58.3%;  Pred. No. 9.5;
  Matches   49;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      16 TGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCCGGCGCGTTGCCG 75
           || |||| || |   | ||    |  | ||||||    |   |||||| |  | |||| |
Db     271 TGCACGAACCTCCAGCCCGAGGTGAGGAAGCCGATGCCCTTGGCGCCCCGGCCNTTGCGG 212

Qy      76 CGCCTCGAGGTGCTGGAGGACCGC 99
           | |||  || |||  | |  ||||
Db     211 CCCCTGAAGCTGCGCGGGCCCCGC 188


RESULT 5
US-09-864-761-12628/c
; Sequence 12628, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12628
;   LENGTH: 531
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL121673.27
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.68
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.75
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.68
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
US-09-864-761-12628

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 531;
  Best Local Similarity   58.1%;  Pred. No. 11;
  Matches   50;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 AGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGC 61
           || ||   ||||  |||| | ||||     |||| |     |||||   ||| || |
Db     388 AGCCCACTCGCCCCTGGAGGCCCCAGTGGGGCGGCCCCCACCAGCCAGGAGATGGGGAAG 329

Qy      62 CCGGCGCGTTGCCGCGCCTCGAGGTG 87
             ||   ||| |||  || |||||||
Db     328 ATGGGCAGTTACCGATCCCCGAGGTG 303


RESULT 6
US-09-933-814-1
; Sequence 1, Application US/09933814
; Patent No. US20020058798A1
; GENERAL INFORMATION:
;  APPLICANT: WALLACH, David
;  APPLICANT:  BOLDIN, Mark
;  APPLICANT:  VARFOLOMEEV, Eugene
;  APPLICANT:  METT, Igor
;  TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APO1 RECEPTORS
;  FILE REFERENCE: WALLACH=16B
;  CURRENT APPLICATION NUMBER: US/09/933,814
;  CURRENT FILING DATE:  2001-08-22
;  PRIOR APPLICATION NUMBER: 08/860,082
;  PRIOR FILING DATE: 1997-08-19
;  PRIOR APPLICATION NUMBER: PCT/US95/16542
;  PRIOR FILING DATE: 1995-12-14
;  PRIOR APPLICATION NUMBER: IL 112022
;  PRIOR FILING DATE: 1994-12-15
;  PRIOR APPLICATION NUMBER: IL 112692
;  PRIOR FILING DATE: 1995-02-19
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 1701
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(768)
US-09-933-814-1

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 1701;
  Best Local Similarity   60.3%;  Pred. No. 9.8;
  Matches   47;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      24 CCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGA 83
           ||  ||| | |   | | || |  |||||| |||||||   | ||   |||||  | |
Db      66 CCTGGCCAGGGCCAGCGAGCCGAGGACAGAGGGCGCGCGGAGGGCCGGGCCGCAGCCCCG 125

Qy      84 GGTGCTGGAGGACCGCGC 101
           |  ||| |  |||| |||
Db     126 GCCGCTTGCAGACCCCGC 143


RESULT 7
US-09-824-134-1
; Sequence 1, Application US/09824134
; Patent No. US20020082401A1
;   GENERAL INFORMATION:
;        APPLICANT: WALLACH, David
;                   BOLDIN, Mark
;                   VARFOLOMEEV, Eugene
;                   METT, Igor
;        TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS/APO1
;                            RECEPTORS
;        NUMBER OF SEQUENCES: 2
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
;             STREET: 419 Seventh Street N.W., Ste. 300
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: United States of America
;             ZIP: 20004
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/824,134
;             FILING DATE: 03-Apr-2001
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/860,082
;             FILING DATE: <Unknown>
;             APPLICATION NUMBER: IL 112022
;             FILING DATE: 15-DEC-1994
;             APPLICATION NUMBER: IL 112692
;             FILING DATE: 19-FEB-1995
;             APPLICATION NUMBER: IL 114615
;             FILING DATE: 16-JUL-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: BROWDY, Roger L.
;             REGISTRATION NUMBER: 25,618
;             REFERENCE/DOCKET NUMBER: WALLACH=16
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (202) 628-5197
;             TELEFAX: (202) 737-3528
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1701 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  1..768
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
```

```
US-09-824-134-1

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 1701;
  Best Local Similarity   60.3%;  Pred. No. 9.8;
  Matches   47;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      24 CCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGA 83
           ||  ||| | |   | | || |  |||||| |||||||   | ||   |||||  | |
Db      66 CCTGGCCAGGGCCAGCGAGCCGAGGACAGAGGGCGCGCGGAGGGCCGGGCCGCAGCCCCG 125

Qy      84 GGTGCTGGAGGACCGCGC 101
           |  ||| |  |||| |||
Db     126 GCCGCTTGCAGACCCCGC 143


RESULT 8
US-09-808-387-29
; Sequence 29, Application US/09808387
; Patent No. US20020132293A1
; GENERAL INFORMATION:
;  APPLICANT: Kaia Palm
;  APPLICANT:  Tonis Timmusk
;  APPLICANT:  CeMines Research
;  TITLE OF INVENTION: MAMMALIAN NEURALIZED FAMILY OF
;  TITLE OF INVENTION:  TRANSCRIPTION REGULATORS AND USES THEREFOR
;  FILE REFERENCE: CEMRES.001A
;  CURRENT APPLICATION NUMBER: US/09/808,387
;  CURRENT FILING DATE:  2001-03-14
;  NUMBER OF SEQ ID NOS: 48
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 789
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-808-387-29

  Query Match             27.7%;  Score 28;  DB 10;  Length 789;
  Best Local Similarity   60.5%;  Pred. No. 13;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      19 ACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGC 78
           | |  | ||||| || | ||| | | || |   || |||| ||||| ||     |||| |
Db     406 AAGGTCAACGCCGGCTGCCGGCTCCTGCAGCGTGAGGGCGTGCCCGTCGGCGCCCCGCTC 465

Qy      79 CTCGAGGTGCTGGAGG 94
              |  ||| |||| |
Db     466 TGGGCCGTGATGGACG 481


RESULT 9
US-09-917-800A-1719/c
; Sequence 1719, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,080
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,536
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1719
;   LENGTH: 1408
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;  OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_024125
US-09-917-800A-1719

  Query Match             27.7%;  Score 28;  DB 10;  Length 1408;
  Best Local Similarity   60.5%;  Pred. No. 13;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      16 TGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCG 75
           | | ||   |  ||| |||| |  | | ||| | | |  |||||| ||    |||| | ||
Db     675 TTGGCGGGCGCGGCCGGCGGCCCCGCGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCG 616

Qy      76 CGCCTCGAGGTGCTGG 91
              | ||||  || | ||
Db     615 GGGCTCGGCGTCCCGG 600


RESULT 10
US-09-815-242-7925
; Sequence 7925, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7925
;   LENGTH: 1446
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1446)
```

```
US-09-815-242-7925

  Query Match             27.7%;  Score 28;  DB 10;  Length 1446;
  Best Local Similarity   58.3%;  Pred. No. 13;
  Matches   49;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      18 GACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCG 77
           ||  | | | |||   || | |  ||||   || ||| | |   ||  || |  |  ||
Db    1215 GATCATCGAGGCCAAGGGCCTGAAGCAGGTCACCGACAGTGGCGCCATCGAGAAGATGCT 1274

Qy      78 CCTCGAGGTGCTGGAGGACCGCGC 101
           |  ||||||||||| || |  |||
Db    1275 CGACGAGGTGCTGGCGGCCAACGC 1298


RESULT 11
US-09-789-836-10/c
; Sequence 10, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
;  APPLICANT: BRIGHAM, KENNETH L.
;  APPLICANT:  STECENKO, ARLENE A.
;  APPLICANT:  SEALY, LINDA
;  TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
;  FILE REFERENCE: N-6977
;  CURRENT APPLICATION NUMBER: US/09/789,836
;  CURRENT FILING DATE:  2001-02-20
;  PRIOR APPLICATION NUMBER: 60/183,584
;  PRIOR FILING DATE: 2000-02-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 1739
;   TYPE: DNA
;   ORGANISM: Rattus sp.
US-09-789-836-10

  Query Match             27.7%;  Score 28;  DB 10;  Length 1739;
  Best Local Similarity   60.5%;  Pred. No. 12;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      16 TGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCG 75
           | | ||  |   ||| |||| |  | | ||| | | |  |||||| ||   |||| | ||
Db    761 TTGGCGGGCGCGGCCGGCGGCCCCGCGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCG 702

Qy      76 CGCCTCGAGGTGCTGG 91
              | ||||  || | ||
Db    701 GGGCTCGGCGTCCCGG 686


RESULT 12
US-09-925-299-166/c
; Sequence 166, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA102
;  CURRENT APPLICATION NUMBER: US/09/925,299
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05883
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1556
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 166
;   LENGTH: 2061
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-299-166

  Query Match             27.7%;  Score 28;  DB 10;  Length 2061;
  Best Local Similarity   59.7%;  Pred. No. 12;
  Matches   43;  Conservative    2;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      27 CGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGT 86
           ||| | || |  |||||| |  ||  |  || |||: |:||     | | ||||| | |
Db      74 CGCGCTCGCTGAGGTGCAGGCTGCACGCACCGGGCCYCGYGCCCGCGCCCTCCTCGGGCT 15

Qy      87 GCTGGAGGACCG 98
           ||||  |  |||
Db      14 GCTGCGGTGCCG 3


RESULT 13
US-09-923-876-1480/c
; Sequence 1480, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 1480
;   LENGTH: 224
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700158889H1
;   NAME/KEY: unsure
;   LOCATION: 96, 116, 148, 173, 189, 211
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-1480

  Query Match             27.5%;  Score 27.8;  DB 10;  Length 224;
  Best Local Similarity   58.8%;  Pred. No. 17;
  Matches   47;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      22 ACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTC 81
           | || || ||    || |||    | |   ||||| ||    | ||    ||| |||| |
Db     107 AGCCGCGTCCTNCATCAGGTTGTACAGCGGGACGGAGCCGGAGCCGGCCAGCCACGCCGC 48

Qy      82 GAGGTGCTGGAGGACCGCGC 101
           ||||| ||| | | || |||
Db      47 GAGGTACTGCACGCCCACGC 28


RESULT 14
US-09-764-887-416
; Sequence 416, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 416
;   LENGTH: 464
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-416
```

Query Match 27.5%; Score 27.8; DB 10; Length 464;
Best Local Similarity 59.5%; Pred. No. 16;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy       3 GACCGAACGCCGATGGACGACCCACGCCCCGGTCGGGTGCAGCCGACAGACGGCGCGCC 62
           ||||   | | |  |   | ||  || ||  ||||  || | |||       | ||||||
Db     117 GACCCCGCCCAGGAGCCGGGCCGGCGGCCTGGGTCTCGTACTGCCTGGCCGGGACGCGCC 176

Qy      63 CGGCGCGTTGCCGCGCCTC 81
             ||| ||| ||| ||||||
Db     177 TGGCTCGTGGCCCCGCCTC 195
```

RESULT 15
US-09-764-887-457
; Sequence 457, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA113
; CURRENT APPLICATION NUMBER: US/09/764,887
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 457
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-887-457

Query Match 27.5%; Score 27.8; DB 10; Length 464;
Best Local Similarity 59.5%; Pred. No. 16;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy       3 GACCGAACGCCGATGGACGACCCACGCCCCGGTCGGGTGCAGCCGACAGACGGCGCGCC 62
           ||||   | | |  |   | ||  || ||  ||||  || | |||       | ||||||
Db     117 GACCCCGCCCAGGAGCCGGGCCGGCGGCCTGGGTCTCGTACTGCCTGGCCGGGACGCGCC 176

Qy      63 CGGCGCGTTGCCGCGCCTC 81
             ||| ||| ||| ||||||
Db     177 TGGCTCGTGGCCCCGCCTC 195
```

Search completed: November 5, 2002, 23:01:42
Job time : 83.2537 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300
Perfect score: 101
Sequence: 1 gagaccgaacgccgatggac.....gaggtgctggaggaccgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	68750	1	AF210843	AF210843 Sorangium
3		101	100.0	68750	6	AR193029	AR193029 Sequence
4		101	100.0	68750	6	AR199551	AR199551 Sequence
5		101	100.0	68750	6	AR199559	AR199559 Sequence
6		101	100.0	68750	6	AR199567	AR199567 Sequence
7		101	100.0	68750	6	AR201097	AR201097 Sequence
8		101	100.0	68750	6	AR208671	AR208671 Sequence
9		101	100.0	71989	6	AR172664	AR172664 Sequence
10		36.8	36.4	80966	2	AC119719	AC119719 Rattus no
11	c	36.6	36.2	155406	2	AC105514	AC105514 Rattus no
12		35.6	35.2	92600	2	AC128756	AC128756 Rattus no
13	c	34.8	34.5	134366	2	AC119661	AC119661 Rattus no
14	c	34.6	34.3	152355	2	AC109839	AC109839 Rattus no
15	c	34.4	34.1	154607	2	AC008193	AC008193 Drosophil
16		34.2	33.9	896	11	PM12F6G	AL684418 Penicilli
17		34	33.7	190459	2	AP005419	AP005419 Oryza sat
18		33.8	33.5	83383	2	AC095331	AC095331 Rattus no
19		33.8	33.5	176695	2	AC126065	AC126065 Rattus no
20		33.6	33.3	2936	1	AF220005	AF220005 Rhodospir
21		33.6	33.3	111084	9	AC006486	AC006486 Homo sapi
22		33.4	33.1	719	9	HSA324403	AJ324403 Homo sapi
23	c	33.4	33.1	1266	1	MFA251337	AJ251337 Mycobacte
24	c	33.4	33.1	69160	2	AC094374	AC094374 Rattus no
25		33.4	33.1	84464	9	HS782D21	AL022399 Human DNA
26	c	33.2	32.9	97967	2	AC103236	AC103236 Rattus no
27		33.2	32.9	181400	2	AC111558	AC111558 Rattus no
28	c	33.2	32.9	185575	2	AC125578	AC125578 Rattus no
29	c	33.2	32.9	262549	2	AC113623	AC113623 Rattus no
30		32.8	32.5	99539	2	AC097740	AC097740 Rattus no
31	c	32.8	32.5	172435	8	OSJN00163	AL662963 Oryza sat
32	c	32.6	32.3	106946	2	AC118298	AC118298 Rattus no
33		32.6	32.3	125481	2	AC120974	AC120974 Rattus no
34		32.6	32.3	143330	2	AC094396	AC094396 Rattus no
35		32.6	32.3	154587	2	AC127744	AC127744 Rattus no
36	c	32.6	32.3	190457	2	AC119326	AC119326 Rattus no
37		32.4	32.1	28731	2	AC121461	AC121461 Rattus no
38		32.4	32.1	76292	2	AC128848	AC128848 Rattus no
39		32.4	32.1	139629	2	AP003809	AP003809 Oryza sat
40	c	32.4	32.1	157448	2	AC131139	AC131139 Rattus no
41	c	32.2	31.9	1156	8	AF342809	AF342809 Zea mays
42		32.2	31.9	14675	1	AE007017	AE007017 Mycobacte
43		32.2	31.9	37750	1	SC6D7	AL133213 Streptomy
44		32.2	31.9	39150	1	MTCY21B4	Z80108 Mycobacteri
45	c	32.2	31.9	46202	2	AC096789	AC096789 Rattus no

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"

CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"

CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"

gene 1998. .6263
/gene="epoA"

CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADREIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKKLKSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"

gene 6260. .10492
/gene="epoB"

CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"

gene 10639. .16137
/gene="epoC"

CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAIPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 7.7e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28200 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 28259

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   28260 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 28300


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
  PUBMED    10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF25907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKPTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFVRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAANLNLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251. .21749
                     /gene="epoB"
     CDS             16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match             100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912
```

```
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a   22456 c   25539 g   11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 7.4e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a   23531 c   26617 g   11731 t      2 others
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28200 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 28259

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||| ||||||||||
Db  28260 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 28300


RESULT 10
AC119719
LOCUS       AC119719               80966 bp    DNA     linear   HTG 23-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-525B13, *** SEQUENCING IN PROGRESS
            ***, 45 unordered pieces.
ACCESSION   AC119719
VERSION     AC119719.4  GI:21902752
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 80966)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 80966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 80966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 18, 2002 this sequence version replaced gi:20467935.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWTP
```

```
    Center clone name: CH230-525B13
-------------- Summary Statistics
    Sequencing vector: Plasmid;
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 31053 bases at least Q40
    Consensus quality: 34998 bases at least Q30
    Consensus quality: 37207 bases at least Q20
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1028: contig of 1028 bp in length
*     1029     1128: gap of unknown length
*     1129     2163: contig of 1035 bp in length
*     2164     2263: gap of unknown length
*     2264     3343: contig of 1080 bp in length
*     3344     3443: gap of unknown length
*     3444     4604: contig of 1161 bp in length
*     4605     4704: gap of unknown length
*     4705     6283: contig of 1579 bp in length
*     6284     6383: gap of unknown length
*     6384     7620: contig of 1237 bp in length
*     7621     7720: gap of unknown length
*     7721     8789: contig of 1069 bp in length
*     8790     8889: gap of unknown length
*     8890    10211: contig of 1322 bp in length
*    10212    10311: gap of unknown length
*    10312    11459: contig of 1148 bp in length
*    11460    11559: gap of unknown length
*    11560    13008: contig of 1449 bp in length
*    13009    13108: gap of unknown length
*    13109    14251: contig of 1143 bp in length
*    14252    14351: gap of unknown length
*    14352    15353: contig of 1002 bp in length
*    15354    15453: gap of unknown length
*    15454    16525: contig of 1072 bp in length
*    16526    16625: gap of unknown length
*    16626    18040: contig of 1415 bp in length
*    18041    18140: gap of unknown length
*    18141    19338: contig of 1198 bp in length
*    19339    19438: gap of unknown length
*    19439    20604: contig of 1166 bp in length
*    20605    20704: gap of unknown length
*    20705    21934: contig of 1230 bp in length
*    21935    22034: gap of unknown length
*    22035    23843: contig of 1809 bp in length
*    23844    23943: gap of unknown length
*    23944    24958: contig of 1015 bp in length
*    24959    25058: gap of unknown length
*    25059    26621: contig of 1563 bp in length
*    26622    26721: gap of unknown length
*    26722    28050: contig of 1329 bp in length
*    28051    28150: gap of unknown length
*    28151    29585: contig of 1435 bp in length
*    29586    29685: gap of unknown length
*    29686    30962: contig of 1277 bp in length
*    30963    31062: gap of unknown length
*    31063    32526: contig of 1464 bp in length
*    32527    32626: gap of unknown length
*    32627    34036: contig of 1410 bp in length
*    34037    34136: gap of unknown length
*    34137    35647: contig of 1511 bp in length
*    35648    35747: gap of unknown length
*    35748    37885: contig of 2138 bp in length
*    37886    37985: gap of unknown length
*    37986    39418: contig of 1433 bp in length
*    39419    39518: gap of unknown length
*    39519    41325: contig of 1807 bp in length
*    41326    41425: gap of unknown length
*    41426    43316: contig of 1891 bp in length
*    43317    43416: gap of unknown length
*    43417    45357: contig of 1941 bp in length
*    45358    45457: gap of unknown length
*    45458    46798: contig of 1341 bp in length
*    46799    46898: gap of unknown length
*    46899    48908: contig of 2010 bp in length
*    48909    49008: gap of unknown length
*    49009    50640: contig of 1632 bp in length
*    50641    50740: gap of unknown length
*    50741    52540: contig of 1800 bp in length
*    52541    52640: gap of unknown length
*    52641    54111: contig of 1471 bp in length
*    54112    54211: gap of unknown length
*    54212    56032: contig of 1821 bp in length
*    56033    56132: gap of unknown length
*    56133    58644: contig of 2512 bp in length
*    58645    58744: gap of unknown length
*    58745    61380: contig of 2636 bp in length
*    61381    61480: gap of unknown length
*    61481    64955: contig of 3475 bp in length
*    64956    65055: gap of unknown length
*    65056    67903: contig of 2848 bp in length
*    67904    68003: gap of unknown length
*    68004    70437: contig of 2434 bp in length
*    70438    70537: gap of unknown length
*    70538    74213: contig of 3676 bp in length
*    74214    74313: gap of unknown length
*    74314    77478: contig of 3165 bp in length
*    77479    77578: gap of unknown length
*    77579    80966: contig of 3388 bp in length.
FEATURES             Location/Qualifiers
     source          1. .80966
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-525B13"
BASE COUNT    22656 a  16980 c  15714 g  21136 t   4480 others
ORIGIN

  Query Match            36.4%;  Score 36.8;  DB 2;  Length 80966;
  Best Local Similarity  63.6%;  Pred. No. 1.3e+02;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||   | ||  ||| |||  ||| | |||||| ||  | |  |||||||||| || |   |
Db   17536 GACCACAAATTCCGCTGGGAGACGCCCGCCCGAGGGGGAGGCCAGCCGACAAACCGGCAG 17595

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGC 88
           ||| |||   || ||||  || || ||
Db   17596 GCCGCCGCCAGGCGGCGCGCCGGGGCGC 17623


RESULT 11
AC105514/c
LOCUS       AC105514              155406 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-183L2, *** SEQUENCING IN PROGRESS
            ***, 69 unordered pieces.
ACCESSION   AC105514
VERSION     AC105514.2  GI:21736378
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 155406)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
```

Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 155406)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 155406)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jul 12, 2002 this sequence version replaced gi:18092736.
---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GNCG
Center clone name: CH230-183L2
---------- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 105305 bases at least Q40
Consensus quality: 112375 bases at least Q30
Consensus quality: 117353 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 69 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	1042: contig of 1042 bp in length
*	1043	1142: gap of unknown length
*	1143	2187: contig of 1045 bp in length
*	2188	2287: gap of unknown length
*	2288	3792: contig of 1505 bp in length
*	3793	3892: gap of unknown length
*	3893	5245: contig of 1353 bp in length
*	5246	5345: gap of unknown length
*	5346	6480: contig of 1135 bp in length
*	6481	6580: gap of unknown length
*	6581	7612: contig of 1032 bp in length
*	7613	7712: gap of unknown length
*	7713	9240: contig of 1528 bp in length
*	9241	9340: gap of unknown length
*	9341	11080: contig of 1740 bp in length
*	11081	11180: gap of unknown length
*	11181	12843: contig of 1663 bp in length
*	12844	12943: gap of unknown length
*	12944	14336: contig of 1393 bp in length
*	14337	14436: gap of unknown length
*	14437	15473: contig of 1037 bp in length
*	15474	15573: gap of unknown length
*	15574	16609: contig of 1036 bp in length
*	16610	16709: gap of unknown length
*	16710	17799: contig of 1090 bp in length
*	17800	17899: gap of unknown length
*	17900	20092: contig of 2193 bp in length
*	20093	20192: gap of unknown length
*	20193	21535: contig of 1343 bp in length
*	21536	21635: gap of unknown length
*	21636	22762: contig of 1127 bp in length
*	22763	22862: gap of unknown length
*	22863	24615: contig of 1753 bp in length
*	24616	24715: gap of unknown length
*	24716	26443: contig of 1728 bp in length
*	26444	26543: gap of unknown length
*	26544	28712: contig of 2169 bp in length
*	28713	28812: gap of unknown length
*	28813	30660: contig of 1848 bp in length
*	30661	30760: gap of unknown length
*	30761	31997: contig of 1237 bp in length
*	31998	32097: gap of unknown length
*	32098	33863: contig of 1766 bp in length
*	33864	33963: gap of unknown length
*	33964	36028: contig of 2065 bp in length
*	36029	36128: gap of unknown length
*	36129	37499: contig of 1371 bp in length
*	37500	37599: gap of unknown length
*	37600	39065: contig of 1466 bp in length
*	39066	39165: gap of unknown length
*	39166	40476: contig of 1311 bp in length
*	40477	40576: gap of unknown length
*	40577	41785: contig of 1209 bp in length
*	41786	41885: gap of unknown length
*	41886	43400: contig of 1515 bp in length
*	43401	43500: gap of unknown length
*	43501	45175: contig of 1675 bp in length
*	45176	45275: gap of unknown length
*	45276	46499: contig of 1224 bp in length
*	46500	46599: gap of unknown length
*	46600	48151: contig of 1552 bp in length
*	48152	48251: gap of unknown length
*	48252	49888: contig of 1637 bp in length
*	49889	49988: gap of unknown length
*	49989	51865: contig of 1877 bp in length
*	51866	51965: gap of unknown length
*	51966	54365: contig of 2400 bp in length
*	54366	54465: gap of unknown length

```
            *    54466    56114: contig of 1649 bp in length
            *    56115    56214: gap of unknown length
            *    56215    57910: contig of 1696 bp in length
            *    57911    58010: gap of unknown length
            *    58011    59278: contig of 1268 bp in length
            *    59279    59378: gap of unknown length
            *    59379    61050: contig of 1672 bp in length
            *    61051    61150: gap of unknown length
            *    61151    63345: contig of 2195 bp in length
            *    63346    63445: gap of unknown length
            *    63446    65431: contig of 1986 bp in length
            *    65432    65531: gap of unknown length
            *    65532    67268: contig of 1737 bp in length
            *    67269    67368: gap of unknown length
            *    67369    69597: contig of 2229 bp in length
            *    69598    69697: gap of unknown length
            *    69698    71840: contig of 2143 bp in length
            *    71841    71940: gap of unknown length
            *    71941    73515: contig of 1575 bp in length
            *    73516    73615: gap of unknown length
            *    73616    75334: contig of 1719 bp in length
            *    75335    75434: gap of unknown length
            *    75435    77060: contig of 1626 bp in length
            *    77061    77160: gap of unknown length
            *    77161    79469: contig of 2309 bp in length
            *    79470    79569: gap of unknown length
            *    79570    82104: contig of 2535 bp in length
            *    82105    82204: gap of unknown length
            *    82205    83687: contig of 1483 bp in length
            *    83688    83787: gap of unknown length
            *    83788    87594: contig of 3807 bp in length
            *    87595    87694: gap of unknown length
            *    87695    89779: contig of 2085 bp in length
            *    89780    89879: gap of unknown length
            *    89880    92763: contig of 2884 bp in length
            *    92764    92863: gap of unknown length
            *    92864    95573: contig of 2710 bp in length
            *    95574    95673: gap of unknown length
            *    95674    97554: contig of 1881 bp in length

  Query Match             36.2%;  Score 36.6;  DB 2;  Length 155406;
  Best Local Similarity   60.6%;  Pred. No. 1.2e+02;
  Matches   60;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       3 GACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCC 62
           | ||   ||||| | | ||  || ||| ||| | ||||     ||| | ||||||||||
Db   38961 GGCCCCCCGCCGCTCGCCGCGCCCCGCGCGCCGCCGGGCCGCCCCGCCCGACGGCGCGCC 38902

Qy      63 CGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
            ||||||   || |  | |  |  || |  ||  | | |
Db   38901 GGGCGCGCCCCCCCCGCGCCCGACGCCGCCGGCGCTCCC 38863


RESULT 12
AC128756
LOCUS       AC128756               92600 bp    DNA     linear   HTG 24-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-187J18, *** SEQUENCING IN PROGRESS
            ***, 55 unordered pieces.
ACCESSION   AC128756
VERSION     AC128756.1  GI:21952558
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 92600)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92600)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
               Center: Baylor College of Medicine
               Center code: BCM
               Web site: http://www.hgsc.bcm.tmc.edu/
               Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
               Center project name: KALL
               Center clone name: CH230-187J18
            -------------- Summary Statistics
               Sequencing vector: Plasmid;
               Chemistry: Dye-terminator Big Dye: 100% of reads
               Assembly program: Phrap; version 0.990329
               Consensus quality: 31018 bases at least Q40
               Consensus quality: 33605 bases at least Q30
               Consensus quality: 35249 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 55 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1078: contig of 1078 bp in length
            *     1079     1178: gap of unknown length
            *     1179     2294: contig of 1116 bp in length
            *     2295     2394: gap of unknown length
            *     2395     3433: contig of 1039 bp in length
```

```
*     3434     3533: gap of unknown length
*     3534     4578: contig of 1045 bp in length
*     4579     4678: gap of unknown length
*     4679     5846: contig of 1168 bp in length
*     5847     5946: gap of unknown length
*     5947     6975: contig of 1029 bp in length
*     6976     7075: gap of unknown length
*     7076     8196: contig of 1121 bp in length
*     8197     8296: gap of unknown length
*     8297     9377: contig of 1081 bp in length
*     9378     9477: gap of unknown length
*     9478    10527: contig of 1050 bp in length
*    10528    10627: gap of unknown length
*    10628    11655: contig of 1028 bp in length
*    11656    11755: gap of unknown length
*    11756    12922: contig of 1167 bp in length
*    12923    13022: gap of unknown length
*    13023    14141: contig of 1119 bp in length
*    14142    14241: gap of unknown length
*    14242    15780: contig of 1539 bp in length
*    15781    15880: gap of unknown length
*    15881    17348: contig of 1468 bp in length
*    17349    17448: gap of unknown length
*    17449    19154: contig of 1706 bp in length
*    19155    19254: gap of unknown length
*    19255    20371: contig of 1117 bp in length
*    20372    20471: gap of unknown length
*    20472    21980: contig of 1509 bp in length
*    21981    22080: gap of unknown length
*    22081    23513: contig of 1433 bp in length
*    23514    23613: gap of unknown length
*    23614    24728: contig of 1115 bp in length
*    24729    24828: gap of unknown length
*    24829    26482: contig of 1654 bp in length
*    26483    26582: gap of unknown length
*    26583    28567: contig of 1985 bp in length
*    28568    28667: gap of unknown length
*    28668    30067: contig of 1400 bp in length
*    30068    30167: gap of unknown length
*    30168    31529: contig of 1362 bp in length
*    31530    31629: gap of unknown length
*    31630    33120: contig of 1491 bp in length
*    33121    33220: gap of unknown length
*    33221    34839: contig of 1619 bp in length
*    34840    34939: gap of unknown length
*    34940    36666: contig of 1727 bp in length
*    36667    36766: gap of unknown length
*    36767    38267: contig of 1501 bp in length
*    38268    38367: gap of unknown length
*    38368    40109: contig of 1742 bp in length
*    40110    40209: gap of unknown length
*    40210    41984: contig of 1775 bp in length
*    41985    42084: gap of unknown length
*    42085    43627: contig of 1543 bp in length
*    43628    43727: gap of unknown length
*    43728    45130: contig of 1403 bp in length
*    45131    45230: gap of unknown length
*    45231    46973: contig of 1743 bp in length
*    46974    47073: gap of unknown length
*    47074    48323: contig of 1250 bp in length
*    48324    48423: gap of unknown length
*    48424    50788: contig of 2365 bp in length
*    50789    50888: gap of unknown length
*    50889    53130: contig of 2242 bp in length
*    53131    53230: gap of unknown length
*    53231    54419: contig of 1189 bp in length
*    54420    54519: gap of unknown length
*    54520    55819: contig of 1300 bp in length
*    55820    55919: gap of unknown length
*    55920    57191: contig of 1272 bp in length
*    57192    57291: gap of unknown length
*    57292    59542: contig of 2251 bp in length
*    59543    59642: gap of unknown length
*    59643    61685: contig of 2043 bp in length
*    61686    61785: gap of unknown length
*    61786    64099: contig of 2314 bp in length
*    64100    64199: gap of unknown length
*    64200    66339: contig of 2140 bp in length
*    66340    66439: gap of unknown length
*    66440    67729: contig of 1290 bp in length
*    67730    67829: gap of unknown length
*    67830    69978: contig of 2149 bp in length
*    69979    70078: gap of unknown length
*    70079    72142: contig of 2064 bp in length
*    72143    72242: gap of unknown length
*    72243    73783: contig of 1541 bp in length
*    73784    73883: gap of unknown length
*    73884    75646: contig of 1763 bp in length
*    75647    75746: gap of unknown length
*    75747    77543: contig of 1797 bp in length
*    77544    77643: gap of unknown length
*    77644    78857: contig of 1214 bp in length
*    78858    78957: gap of unknown length
*    78958    80450: contig of 1493 bp in length
*    80451    80550: gap of unknown length
*    80551    83143: contig of 2593 bp in length
*    83144    83243: gap of unknown length
*    83244    85335: contig of 2092 bp in length
*    85336    85435: gap of unknown length
*    85436    87511: contig of 2076 bp in length
*    87512    87611: gap of unknown length
*    87612    89843: contig of 2232 bp in length
*    89844    89943: gap of unknown length
*    89944    92600: contig of 2657 bp in length.
FEATURES             Location/Qualifiers
     source          1. .92600
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-187J18"

  Query Match             35.2%;  Score 35.6;  DB 2;  Length 92600;
  Best Local Similarity   62.2%;  Pred. No. 2.2e+02;
  Matches   56;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       6 CGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGG 65
           ||   || ||  | |  ||  ||| ||||| |||  | |||  || |  ||||| | |||
Db     868 CGGGAGCGGACAGGCAGCCAGCGCGCGCGGACGGAGCACACGGCAGGCGCACGCGG 927

Qy      66 CGCGTTGCCGCGCCCTCGAGGTGCTGGAGGA 95
           | ||  | | | || | ||| || ||| ||
Db     928 CTCGACGACCCCCCGCAAGGAGCCGGACGA 957


RESULT 13
AC119661/c
LOCUS       AC119661              134366 bp    DNA     linear   HTG 23-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-256E2, *** SEQUENCING IN PROGRESS
            ***, 57 unordered pieces.
ACCESSION   AC119661
VERSION     AC119661.5  GI:21902745
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 134366)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
```

```
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134366)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 134366)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 18, 2002 this sequence version replaced gi:20467932.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GWSV
                Center clone name: CH230-256E2
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 83294 bases at least Q40
                Consensus quality: 87027 bases at least Q30
                Consensus quality: 90347 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 57 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1237: contig of 1237 bp in length
            *     1238     1337: gap of unknown length
            *     1338     2424: contig of 1087 bp in length
            *     2425     2524: gap of unknown length
            *     2525     3669: contig of 1145 bp in length
            *     3670     3769: gap of unknown length
            *     3770     4928: contig of 1159 bp in length
            *     4929     5028: gap of unknown length
            *     5029     6169: contig of 1141 bp in length
            *     6170     6269: gap of unknown length
            *     6270     7515: contig of 1246 bp in length
            *     7516     7615: gap of unknown length
            *     7616     8717: contig of 1102 bp in length
            *     8718     8817: gap of unknown length
            *     8818     9940: contig of 1123 bp in length
            *     9941    10040: gap of unknown length
            *    10041    11081: contig of 1041 bp in length
            *    11082    11181: gap of unknown length
            *    11182    12261: contig of 1080 bp in length
            *    12262    12361: gap of unknown length
            *    12362    13506: contig of 1145 bp in length
            *    13507    13606: gap of unknown length
            *    13607    15106: contig of 1500 bp in length
            *    15107    15206: gap of unknown length
            *    15207    16380: contig of 1174 bp in length
            *    16381    16480: gap of unknown length
            *    16481    17966: contig of 1486 bp in length
            *    17967    18066: gap of unknown length
            *    18067    19105: contig of 1039 bp in length
            *    19106    19205: gap of unknown length
            *    19206    20568: contig of 1363 bp in length
            *    20569    20668: gap of unknown length
            *    20669    21680: contig of 1012 bp in length
            *    21681    21780: gap of unknown length
            *    21781    22947: contig of 1167 bp in length
            *    22948    23047: gap of unknown length
            *    23048    24154: contig of 1107 bp in length
            *    24155    24254: gap of unknown length
            *    24255    25444: contig of 1190 bp in length
            *    25445    25544: gap of unknown length
            *    25545    27949: contig of 2405 bp in length
            *    27950    28049: gap of unknown length
            *    28050    29450: contig of 1401 bp in length
            *    29451    29550: gap of unknown length
            *    29551    30746: contig of 1196 bp in length
            *    30747    30846: gap of unknown length
            *    30847    32179: contig of 1333 bp in length
            *    32180    32279: gap of unknown length
            *    32280    33614: contig of 1335 bp in length
            *    33615    33714: gap of unknown length
            *    33715    34960: contig of 1246 bp in length
            *    34961    35060: gap of unknown length
            *    35061    37362: contig of 2302 bp in length
            *    37363    37462: gap of unknown length
            *    37463    39351: contig of 1889 bp in length
            *    39352    39451: gap of unknown length
            *    39452    40658: contig of 1207 bp in length
            *    40659    40758: gap of unknown length
            *    40759    42160: contig of 1402 bp in length
            *    42161    42260: gap of unknown length
            *    42261    44051: contig of 1791 bp in length
            *    44052    44151: gap of unknown length
            *    44152    45525: contig of 1374 bp in length
            *    45526    45625: gap of unknown length
            *    45626    47255: contig of 1630 bp in length
            *    47256    47355: gap of unknown length
            *    47356    49554: contig of 2199 bp in length
            *    49555    49654: gap of unknown length
            *    49655    51956: contig of 2302 bp in length
            *    51957    52056: gap of unknown length
            *    52057    53921: contig of 1865 bp in length
            *    53922    54021: gap of unknown length
            *    54022    55835: contig of 1814 bp in length
            *    55836    55935: gap of unknown length
```

```
            *    55936    58618: contig of 2683 bp in length
            *    58619    58718: gap of unknown length
            *    58719    61717: contig of 2999 bp in length
            *    61718    61817: gap of unknown length
            *    61818    63684: contig of 1867 bp in length
            *    63685    63784: gap of unknown length
            *    63785    65948: contig of 2164 bp in length
            *    65949    66048: gap of unknown length
            *    66049    67821: contig of 1773 bp in length
            *    67822    67921: gap of unknown length
            *    67922    70305: contig of 2384 bp in length
            *    70306    70405: gap of unknown length
            *    70406    72600: contig of 2195 bp in length
            *    72601    72700: gap of unknown length
            *    72701    74582: contig of 1882 bp in length
            *    74583    74682: gap of unknown length
            *    74683    78292: contig of 3610 bp in length
            *    78293    78392: gap of unknown length
            *    78393    79935: contig of 1543 bp in length
            *    79936    80035: gap of unknown length
            *    80036    83750: contig of 3715 bp in length
            *    83751    83850: gap of unknown length
            *    83851    87621: contig of 3771 bp in length
            *    87622    87721: gap of unknown length
            *    87722    92168: contig of 4447 bp in length
            *    92169    92268: gap of unknown length
            *    92269    96262: contig of 3994 bp in length
            *    96263    96362: gap of unknown length
            *    96363    99910: contig of 3548 bp in length
            *    99911   100010: gap of unknown length
            *   100011   104092: contig of 4082 bp in length
            *   104093   104192: gap of unknown length
            *   104193   110552: contig of 6360 bp in length

  Query Match             34.5%;  Score 34.8;  DB 2;  Length 134366;
  Best Local Similarity   62.8%;  Pred. No. 2.9e+02;
  Matches    54;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      11 GCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGT 70
           |||| ||| || |||   ||||||| ||||    |||| |   ||| ||  ||||||||
Db  49280 GCCGGTGGCCGCCCCTTCCCCGCGGCCGGGGCGCGCCGCCGCGCGGGGCCGCCGGCGCGC 49221

Qy      71 TGCCGCGCCTCGAGGTGCTGGAGGAC 96
           ||   ||  || || |   | || |
Db  49220 GGCGCGGCGCCGCGGGGGGCGCGGGC 49195


RESULT 14
AC109839/c
LOCUS       AC109839               152355 bp    DNA     linear   HTG 17-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-271M10, *** SEQUENCING IN PROGRESS
            ***, 55 unordered pieces.
ACCESSION   AC109839
VERSION     AC109839.3  GI:21743805
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 152355)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,E.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152355)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 152355)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 13, 2002 this sequence version replaced gi:18846423.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GOHA
                Center clone name: CH230-271M10
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 113301 bases at least Q40
                Consensus quality: 117622 bases at least Q30
                Consensus quality: 120952 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 55 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1040: contig of 1040 bp in length
            *     1041     1140: gap of unknown length
            *     1141     2768: contig of 1628 bp in length
            *     2769     2868: gap of unknown length
```

```
*     2869     4091: contig of 1223 bp in length
*     4092     4191: gap of unknown length
*     4192     5976: contig of 1785 bp in length
*     5977     6076: gap of unknown length
*     6077     7167: contig of 1091 bp in length
*     7168     7267: gap of unknown length
*     7268     8549: contig of 1282 bp in length
*     8550     8649: gap of unknown length
*     8650    10359: contig of 1710 bp in length
*    10360    10459: gap of unknown length
*    10460    12098: contig of 1639 bp in length
*    12099    12198: gap of unknown length
*    12199    13472: contig of 1274 bp in length
*    13473    13572: gap of unknown length
*    13573    15615: contig of 2043 bp in length
*    15616    15715: gap of unknown length
*    15716    16923: contig of 1208 bp in length
*    16924    17023: gap of unknown length
*    17024    18314: contig of 1291 bp in length
*    18315    18414: gap of unknown length
*    18415    19430: contig of 1016 bp in length
*    19431    19530: gap of unknown length
*    19531    21598: contig of 2068 bp in length
*    21599    21698: gap of unknown length
*    21699    22722: contig of 1024 bp in length
*    22723    22822: gap of unknown length
*    22823    24919: contig of 2097 bp in length
*    24920    25019: gap of unknown length
*    25020    26524: contig of 1505 bp in length
*    26525    26624: gap of unknown length
*    26625    28975: contig of 2351 bp in length
*    28976    29075: gap of unknown length
*    29076    30712: contig of 1637 bp in length
*    30713    30812: gap of unknown length
*    30813    33147: contig of 2335 bp in length
*    33148    33247: gap of unknown length
*    33248    34748: contig of 1501 bp in length
*    34749    34848: gap of unknown length
*    34849    36700: contig of 1852 bp in length
*    36701    36800: gap of unknown length
*    36801    39406: contig of 2606 bp in length
*    39407    39506: gap of unknown length
*    39507    41430: contig of 1924 bp in length
*    41431    41530: gap of unknown length
*    41531    43892: contig of 2362 bp in length
*    43893    43992: gap of unknown length
*    43993    46871: contig of 2879 bp in length
*    46872    46971: gap of unknown length
*    46972    48525: contig of 1554 bp in length
*    48526    48625: gap of unknown length
*    48626    51785: contig of 3160 bp in length
*    51786    51885: gap of unknown length
*    51886    53510: contig of 1625 bp in length
*    53511    53610: gap of unknown length
*    53611    55888: contig of 2278 bp in length
*    55889    55988: gap of unknown length
*    55989    58034: contig of 2046 bp in length
*    58035    58134: gap of unknown length
*    58135    59998: contig of 1864 bp in length
*    59999    60098: gap of unknown length
*    60099    62273: contig of 2175 bp in length
*    62274    62373: gap of unknown length
*    62374    64829: contig of 2456 bp in length
*    64830    64929: gap of unknown length
*    64930    67734: contig of 2805 bp in length
*    67735    67834: gap of unknown length
*    67835    70831: contig of 2997 bp in length
*    70832    70931: gap of unknown length
*    70932    73720: contig of 2789 bp in length
*    73721    73820: gap of unknown length
*    73821    75955: contig of 2135 bp in length
*    75956    76055: gap of unknown length
*    76056    79308: contig of 3253 bp in length
*    79309    79408: gap of unknown length
*    79409    82621: contig of 3213 bp in length
*    82622    82721: gap of unknown length
*    82722    85703: contig of 2982 bp in length
*    85704    85803: gap of unknown length
*    85804    88896: contig of 3093 bp in length
*    88897    88996: gap of unknown length
*    88997    92297: contig of 3301 bp in length
*    92298    92397: gap of unknown length
*    92398    95549: contig of 3152 bp in length
*    95550    95649: gap of unknown length
*    95650    99040: contig of 3391 bp in length
*    99041    99140: gap of unknown length
*    99141   103190: contig of 4050 bp in length
*   103191   103290: gap of unknown length
*   103291   107171: contig of 3881 bp in length
*   107172   107271: gap of unknown length
*   107272   111292: contig of 4021 bp in length
*   111293   111392: gap of unknown length
*   111393   116175: contig of 4783 bp in length
*   116176   116275: gap of unknown length
*   116276   119026: contig of 2751 bp in length
*   119027   119126: gap of unknown length
*   119127   124826: contig of 5700 bp in length
*   124827   124926: gap of unknown length
*   124927   130259: contig of 5333 bp in length
*   130260   130359: gap of unknown length
*   130360   137587: contig of 7228 bp in length
*   137588   137687: gap of unknown length
*   137688   141811: contig of 4124 bp in length

  Query Match             34.3%;  Score 34.6;  DB 2;  Length 152355;
  Best Local Similarity   67.1%;  Pred. No. 3e+02;
  Matches   49;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       9 ACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGC 68
           ||| ||   |  | |||| | |||||| || | | ||||||   | | ||||||| |  |
Db    4538 ACGGCGCCCGCGGCCCCAAGGCCGCGGGCGCGGGGAGCCGAGGCAAGCCGCGCCCCGGAC 4479

Qy      69 GTTGCCGCGCCTC 81
           |  |||||||| |
Db    4478 GCAGCCGCGCCCC 4466


RESULT 15
AC008193/c
LOCUS       AC008193              154607 bp    DNA     linear   HTG 15-MAR-2002
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR02H01 (D735) RPCI-98
            02.H.1 map 94D-94D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 94 unordered pieces.
ACCESSION   AC008193
VERSION     AC008193.10  GI:19482329
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 154607)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
            Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
            Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,
            Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
            Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154607)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
```

Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

TITLE Direct Submission

JOURNAL Submitted (29-JUL-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

REFERENCE 3 (bases 1 to 154607)

AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.

TITLE Direct Submission

JOURNAL Submitted (15-MAR-2002) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT On Mar 15, 2002 this sequence version replaced gi:6957934.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 94 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	403	contig of 403 bp in length
*	404	483	gap of unknown length
*	484	1060	contig of 577 bp in length
*	1061	1140	gap of unknown length
*	1141	1703	contig of 563 bp in length
*	1704	1783	gap of unknown length
*	1784	2222	contig of 439 bp in length
*	2223	2302	gap of unknown length
*	2303	3069	contig of 767 bp in length
*	3070	3149	gap of unknown length
*	3150	3577	contig of 428 bp in length
*	3578	3657	gap of unknown length
*	3658	4406	contig of 749 bp in length
*	4407	4486	gap of unknown length
*	4487	5644	contig of 1158 bp in length
*	5645	5724	gap of unknown length
*	5725	6939	contig of 1215 bp in length
*	6940	7019	gap of unknown length
*	7020	7779	contig of 760 bp in length
*	7780	7859	gap of unknown length
*	7860	8668	contig of 809 bp in length
*	8669	8748	gap of unknown length
*	8749	9714	contig of 966 bp in length
*	9715	9794	gap of unknown length
*	9795	10377	contig of 583 bp in length
*	10378	10457	gap of unknown length
*	10458	11237	contig of 780 bp in length
*	11238	11317	gap of unknown length
*	11318	12105	contig of 788 bp in length
*	12106	12185	gap of unknown length
*	12186	12976	contig of 791 bp in length
*	12977	13056	gap of unknown length
*	13057	14322	contig of 1266 bp in length
*	14323	14402	gap of unknown length
*	14403	15204	contig of 802 bp in length
*	15205	15284	gap of unknown length
*	15285	16391	contig of 1107 bp in length
*	16392	16471	gap of unknown length
*	16472	17302	contig of 831 bp in length
*	17303	17382	gap of unknown length
*	17383	18798	contig of 1416 bp in length
*	18799	18878	gap of unknown length
*	18879	20188	contig of 1310 bp in length
*	20189	20268	gap of unknown length
*	20269	21661	contig of 1393 bp in length
*	21662	21741	gap of unknown length
*	21742	22795	contig of 1054 bp in length
*	22796	22875	gap of unknown length
*	22876	23922	contig of 1047 bp in length
*	23923	24002	gap of unknown length
*	24003	25192	contig of 1190 bp in length
*	25193	25272	gap of unknown length
*	25273	26693	contig of 1421 bp in length
*	26694	26773	gap of unknown length
*	26774	27574	contig of 801 bp in length
*	27575	27654	gap of unknown length
*	27655	28564	contig of 910 bp in length
*	28565	28644	gap of unknown length
*	28645	30049	contig of 1405 bp in length
*	30050	30129	gap of unknown length
*	30130	31675	contig of 1546 bp in length
*	31676	31755	gap of unknown length
*	31756	33381	contig of 1626 bp in length
*	33382	33461	gap of unknown length
*	33462	34951	contig of 1490 bp in length
*	34952	35031	gap of unknown length
*	35032	36163	contig of 1132 bp in length
*	36164	36243	gap of unknown length
*	36244	38078	contig of 1835 bp in length
*	38079	38158	gap of unknown length
*	38159	39953	contig of 1795 bp in length
*	39954	40033	gap of unknown length
*	40034	40982	contig of 949 bp in length
*	40983	41062	gap of unknown length
*	41063	43012	contig of 1950 bp in length
*	43013	43092	gap of unknown length
*	43093	44793	contig of 1701 bp in length
*	44794	44873	gap of unknown length
*	44874	46731	contig of 1858 bp in length
*	46732	46811	gap of unknown length
*	46812	48752	contig of 1941 bp in length
*	48753	48832	gap of unknown length
*	48833	50461	contig of 1629 bp in length
*	50462	50541	gap of unknown length
*	50542	52858	contig of 2317 bp in length
*	52859	52938	gap of unknown length
*	52939	54831	contig of 1893 bp in length
*	54832	54911	gap of unknown length
*	54912	56618	contig of 1707 bp in length
*	56619	56698	gap of unknown length
*	56699	57856	contig of 1158 bp in length
*	57857	57936	gap of unknown length
*	57937	60795	contig of 2859 bp in length
*	60796	60875	gap of unknown length
*	60876	62728	contig of 1853 bp in length
*	62729	62808	gap of unknown length
*	62809	65904	contig of 3096 bp in length
*	65905	65984	gap of unknown length
*	65985	68416	contig of 2432 bp in length
*	68417	68496	gap of unknown length
*	68497	72254	contig of 3758 bp in length
*	72255	72334	gap of unknown length
*	72335	77110	contig of 4776 bp in length
*	77111	77190	gap of unknown length
*	77191	81641	contig of 4451 bp in length
*	81642	81721	gap of unknown length
*	81722	85348	contig of 3627 bp in length
*	85349	85428	gap of unknown length
*	85429	86976	contig of 1548 bp in length
*	86977	87056	gap of unknown length
*	87057	92459	contig of 5403 bp in length

```
            *    92460    92539: gap of unknown length
            *    92540    99116: contig of 6577 bp in length
            *    99117    99196: gap of unknown length
            *    99197   104719: contig of 5523 bp in length
            *   104720   104799: gap of unknown length
            *   104800   111051: contig of 6252 bp in length
            *   111052   111131: gap of unknown length
            *   111132   114739: contig of 3608 bp in length
            *   114740   114819: gap of unknown length
            *   114820   117464: contig of 2645 bp in length
            *   117465   117544: gap of unknown length
            *   117545   125411: contig of 7867 bp in length
            *   125412   125491: gap of unknown length
            *   125492   132573: contig of 7082 bp in length
            *   132574   132653: gap of unknown length
            *   132654   133260: contig of 607 bp in length
            *   133261   133340: gap of unknown length
            *   133341   134073: contig of 733 bp in length
            *   134074   134153: gap of unknown length
            *   134154   135094: contig of 941 bp in length
            *   135095   135174: gap of unknown length
            *   135175   135921: contig of 747 bp in length
            *   135922   136001: gap of unknown length
            *   136002   136693: contig of 692 bp in length
            *   136694   136773: gap of unknown length
            *   136774   137410: contig of 637 bp in length
            *   137411   137490: gap of unknown length

  Query Match             34.1%;  Score 34.4;  DB 2;  Length 154607;
  Best Local Similarity   60.9%;  Pred. No. 3.3e+02;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      10 CGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCG 69
           |||    || || ||  ||||||| | |||| |  | ||   | ||||||| |  ||||
Db 152836 CGCGCGGGGCCGGCCGCCGCCCGCCGGCGGGGGGCGGCGCGGGGCGGCGCGGCGCGCCGCC 152777

Qy      70 TTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           | |||  | || || || |     ||||||
Db 152776 CGGGCGCCGCCCGGGGCGCCGCCCCCCCGCGC 152745


Search completed: November  5, 2002, 12:33:47
Job time : 717.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300
Perfect score: 101
Sequence: 1 gagaccgaacgccgatggac.....gaggtgctggaggaccgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	31	30.7	3381	4	US-09-009-119-1	Sequence 1, Appli
	10	31	30.7	3381	4	US-09-371-507-1	Sequence 1, Appli
c	11	30.4	30.1	285	6	5244792-8	Patent No. 5244792
c	12	29.8	29.5	23673	4	US-09-773-816-1	Sequence 1, Appli
	13	29.8	29.5	33529	4	US-09-144-085-3	Sequence 3, Appli
c	14	29.6	29.3	1500	4	US-09-593-711A-10	Sequence 10, Appl
c	15	29	28.7	1015	4	US-09-372-422A-31	Sequence 31, Appl
c	16	28.6	28.3	2588	2	US-08-796-414B-6	Sequence 6, Appli
c	17	28.4	28.1	705	4	US-08-998-416-1059	Sequence 1059, Ap
	18	28.4	28.1	1701	3	US-09-357-072-1	Sequence 1, Appli
	19	28.4	28.1	1701	4	US-08-983-502-1	Sequence 1, Appli
	20	28.4	28.1	1701	5	PCT-US95-16542-1	Sequence 1, Appli
	21	28.4	28.1	1701	5	PCT-US96-10521-1	Sequence 1, Appli
	22	28.2	27.9	1723	4	US-09-199-637A-135	Sequence 135, App
c	23	28	27.7	612	4	US-09-484-970B-80	Sequence 80, Appl
	24	28	27.7	3802	1	US-08-404-354B-2	Sequence 2, Appli
	25	28	27.7	3802	1	US-08-314-083B-2	Sequence 2, Appli
	26	28	27.7	3802	1	US-08-435-675B-2	Sequence 2, Appli
	27	28	27.7	3802	1	US-08-336-257A-4	Sequence 4, Appli
	28	28	27.7	3802	3	US-08-884-599-2	Sequence 2, Appli
	29	28	27.7	3802	6	5386025-7	Patent No. 5386025
	30	28	27.7	4464	2	US-08-400-159-7	Sequence 7, Appli
	31	28	27.7	4483	3	US-08-611-729A-7	Sequence 7, Appli
	32	27.8	27.5	3383	5	PCT-US95-09098-1	Sequence 1, Appli
c	33	27.8	27.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	34	27.8	27.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	35	27.6	27.3	755	4	US-08-998-416-748	Sequence 748, App
	36	27.4	27.1	13842	4	US-09-105-537-30	Sequence 30, Appl
	37	27.4	27.1	36778	4	US-09-105-537-5	Sequence 5, Appli
	38	27.4	27.1	38506	3	US-09-320-878-19	Sequence 19, Appl
	39	27.4	27.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	40	27.2	26.9	1550	2	US-08-609-443B-17	Sequence 17, Appl
c	41	27.2	26.9	1550	4	US-08-851-896-17	Sequence 17, Appl
c	42	27	26.7	554	4	US-09-437-054A-1	Sequence 1, Appli
c	43	27	26.7	3472	6	5244792-2	Patent No. 5244792
c	44	27	26.7	5020	1	US-08-459-586-5	Sequence 5, Appli
c	45	27	26.7	5020	2	US-08-282-696-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 4.3e-18;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           ||||||||||||||||||||||||||| ||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           ||||||||||||||||||||||||||| ||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           ||||||||||||||||||||||||||||||||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           ||||||||||||||||||||||||||||||||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  28200 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 28259

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  28260 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 28300


RESULT 9
US-09-009-119-1
; Sequence 1, Application US/09009119
; Patent No. 6160206
; GENERAL INFORMATION:
;  APPLICANT: SATO, Ryo
;  APPLICANT: BOYNTON, John E.
;  APPLICANT: GILLHAM, Nicholas W.
;  APPLICANT: HARRIS, Elizabeth H.
;  TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
;  FILE REFERENCE: substitute sequence listing
;  CURRENT APPLICATION NUMBER: US/09/009,119
;  CURRENT FILING DATE: 1998-01-20
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3381
;   TYPE: DNA
;   ORGANISM: Chlamydomonas reinhardtii
US-09-009-119-1

  Query Match             30.7%;  Score 31;  DB 4;  Length 3381;
  Best Local Similarity   62.0%;  Pred. No. 3.5;
  Matches   49;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy     23 CCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCG 82
          |||||   || |  |||||||||| ||| || | |  || |||  |  | |   |||| |
Db   2982 CCCACCTGCGTGCGCGGGTGCAGCTGACCGAAGCCCGGCACGGACCCGTCCGAGGCCTTG 3041

Qy     83 AGGTGCTGGAGGACCGCGC 101
          | | ||   | || ||||
```

```
Db    3042 CGCTCCTCCCGCACGGCGC 3060


RESULT 10
US-09-371-507-1
; Sequence 1, Application US/09371507
; Patent No. 6346656
; GENERAL INFORMATION:
;  APPLICANT: SATO, Ryo
;  APPLICANT:  BOYNTON, John E.
;  APPLICANT:  GILLHAM, Nicholas W.
;  APPLICANT:  HARRIS, Elizabeth H.
;  TITLE OF INVENTION: Porphyrin Accumulating-Type Herbicide Resistance Gene
;  FILE REFERENCE: substitute sequence listing
;  CURRENT APPLICATION NUMBER: US/09/371,507
;  CURRENT FILING DATE:  1999-08-18
;  PRIOR APPLICATION NUMBER: 09/009,119
;  PRIOR FILING DATE: 1998-01-20
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3381
;   TYPE: DNA
;   ORGANISM: Chlamydomonas reinhardtii
US-09-371-507-1

  Query Match             30.7%;  Score 31;  DB 4;  Length 3381;
  Best Local Similarity   62.0%;  Pred. No. 3.5;
  Matches   49;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      23 CCCACGCCCGCGGTCGGGTGCAGCCGACACGGCGCGCCCGGCGCGTTGCCGCGCCTCG 82
           |||||   || |  |||||||||| ||| || | |  || |||  |  | |   |||| |
Db    2982 CCCACCTGCGTGCGCGGGTGCAGCTGACCGAAGCCCGGCACGGACCCGTCCGAGGCCTTG 3041

Qy      83 AGGTGCTGGAGGACCGCGC 101
           | | ||   | || ||||
Db    3042 CGCTCCTCCCGCACGGCGC 3060


RESULT 11
5244792-8/c
;Patent No. 5244792
;     APPLICANT: BURKE, RAE L.;PACHL, CAROL;VALENZUELA, PABLO D.T.
;     TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPROTEIN
;B FROM HERPES SIMPLEX VIRUS
;     NUMBER OF SEQUENCES: 19
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/587,179
;       FILING DATE: 20-SEP-1990
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 921,730
;       FILING DATE: 20-OCT-1986
;       APPLICATION NUMBER: 597,784
;       FILING DATE: 06-APR-1984
;SEQ ID NO:8:
;       LENGTH: 285
5244792-8

  Query Match             30.1%;  Score 30.4;  DB 6;  Length 285;
  Best Local Similarity   59.1%;  Pred. No. 5.2;
  Matches   52;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      13 CGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCCGGCGCGTTG 72
           ||  ||    |||  || |  ||  |  | | |  | | | ||   | |||| |  || |
Db     173 CGCGGGGGCGCCCTGGCGCATGGCGGACTACGGGGGCCCGTCGTGCCCCCCGTCAGGTAG 114

Qy      73 CCGCGCCTCGAGGTGCTGGAGGACCGCG 100
           ||  |  |||||||||||||||||| |
Db     113 CCTGGGGGCGAGGTGCTGGAGGACCGAG 86


RESULT 12
US-09-773-816-1/c
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
;  APPLICANT: Stanford University
;  APPLICANT:  Khosla, Chaitan
;  TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
;  TITLE OF INVENTION:  ANTAGONISTS
;  FILE REFERENCE: 28600-20210.00
;  CURRENT APPLICATION NUMBER: US/09/773,816
;  CURRENT FILING DATE:  2001-06-08
;  PRIOR APPLICATION NUMBER: US 60/243,458
;  PRIOR FILING DATE: 2000-10-25
;  PRIOR APPLICATION NUMBER: US 60/179,305
;  PRIOR FILING DATE: 2000-01-31
;  NUMBER OF SEQ ID NOS: 1
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 23673
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(23623)
;   OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

  Query Match             29.5%;  Score 29.8;  DB 4;  Length 23673;
  Best Local Similarity   56.7%;  Pred. No. 7;
  Matches   55;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       4 ACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCC 63
           | |  | | ||| | ||| |   || ||||   || |  || |   | |||| |||   |
Db   11442 AGCTCAGGTCGAGGAACGTCTTGCGGCCGCTCACGCGGTCACCGTCCCGACGCCGCCGGC 11383

Qy      64 GGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCG 100
            |||||   |||   ||||   |    ||| ||||||
Db   11382 CGCGCGCACCCGACGCTCGGCCTCGCCGAGCACCGCG 11346


RESULT 13
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             29.5%;  Score 29.8;  DB 4;  Length 33529;
  Best Local Similarity   56.7%;  Pred. No. 7;
  Matches   55;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       3 GACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCC 62
           | |||| |  |    |  | || | | |   |  ||| |||||  ||    ||| |||||
Db   18251 GGCCGAGCCGCTGGTGCTGGCCGAGGGCGCGGCGCGGCTGCAGGTGATGATCGGAGCGCC 18310

Qy      63 CGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGC 99
```

```
          | ||||  |   || | |||||||  |  | || ||
Db  18311 GGACGCGGCGGGCCGGCGCGAGGTGGGGCTGTACAGC 18347


RESULT 14
US-09-593-711A-10/c
; Sequence 10, Application US/09593711A
; Patent No. 6271030
; GENERAL INFORMATION:
;  APPLICANT:    Brett P. Monia
;  APPLICANT:  Madeline M. Butler
;  APPLICANT:  Jacqueline Wyatt
;  TITLE OF INVENTION:    ANTISENSE MODULATION OF C/EBP BETA EXPRESSION
;  FILE REFERENCE:    RTS-0118
;  CURRENT APPLICATION NUMBER: US/09/593,711A
;  CURRENT FILING DATE:  2000-06-14
;  NUMBER OF SEQ ID NOS:    244
; SEQ ID NO 10
;   LENGTH: 1500
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (108)...(998)
US-09-593-711A-10

  Query Match            29.3%;  Score 29.6;  DB 4;  Length 1500;
  Best Local Similarity   61.8%;  Pred. No. 8.1;
  Matches   47;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     16 TGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCG 75
          | | ||  |   ||| |||| |  | | ||| | | |  |||||| ||   |||| | ||
Db    739 TTGGCGGGCGCGGCCGGCGGCCCCGCGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCG 680

Qy     76 CGCCTCGAGGTGCTGG 91
           | ||||  |||| ||
Db    679 GGGCTCGGCGTGCCGG 664


RESULT 15
US-09-372-422A-31/c
; Sequence 31, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
;  APPLICANT: Rudolf Jung
;  APPLICANT:  Francois Barrieu
;  TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
;  FILE REFERENCE: 0919
;  CURRENT APPLICATION NUMBER: US/09/372,422A
;  CURRENT FILING DATE:  1999-08-11
;  PRIOR APPLICATION NUMBER: US 60/098,692
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 1015
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (77)...(863)
US-09-372-422A-31

  Query Match            28.7%;  Score 29;  DB 4;  Length 1015;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
          |||  ||||| |||  || ||     | |||| ||   |  ||||  |     ||||
Db    493 GAGCGCGAACCCCGGCGGCCGCATGCCCCCCGTGGCGAGCCGCAGGAGCAGCGTGGCGGC 434

Qy     61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
          ||||||   || ||  | |||     |  | ||||||||
Db    433 GACGGCGCCCAGCAGCTGCGCGACCCAGTACAAGACCGCGC 393


Search completed: November  5, 2002, 15:43:44
Job time : 147.737 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300
Perfect score: 101
Sequence: 1 gagaccgaacgccgatggac.....gaggtgctggaggaccgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	38.8	38.4	639	10	BE378391	BE378391 601236718
c	2	38.6	38.2	978	12	BF795540	BF795540 602259474
	3	36.4	36.0	1189	17	AG030608	AG030608 Pan trogl
	4	36.2	35.8	888	12	BG809572	BG809572 mgct001xa
c	5	35	34.7	813	17	AG032381	AG032381 Pan trogl
c	6	35	34.7	1020	17	AG063051	AG063051 Pan trogl
	7	34.6	34.3	1059	14	BQ680479	BQ680479 AGENCOURT
	8	34.4	34.1	260	10	AV618417	AV618417 AV618417
	9	34	33.7	983	17	AG082451	AG082451 Pan trogl
	10	33.8	33.5	378	9	AJ449574	AJ449574 AJ449574
	11	33.8	33.5	681	9	AJ452432	AJ452432 AJ452432
c	12	33.6	33.3	724	17	AQ272425	AQ272425 nbxb0027P
	13	33.6	33.3	813	17	AG137902	AG137902 Pan trogl
	14	33.6	33.3	910	17	CNS006ON	AL065629 Drosophil
c	15	33.2	32.9	643	14	BQ788818	BQ788818 WHE4154_B
	16	33	32.7	839	17	CNS004NB	AL054280 Drosophil
c	17	33	32.7	902	13	BI956556	BI956556 HVSMEn000
	18	33	32.7	1105	13	BM474002	BM474002 AGENCOURT
	19	33	32.7	1526	17	AG072563	AG072563 Pan trogl
c	20	32.8	32.5	715	17	BH837262	BH837262 LMCR05001
	21	32.8	32.5	868	17	AG061755	AG061755 Pan trogl
	22	32.8	32.5	1232	17	AG072425	AG072425 Pan trogl
	23	32.8	32.5	1587	17	AG081414	AG081414 Pan trogl
	24	32.6	32.3	734	14	BQ443300	BQ443300 UI-M-EV0-
	25	32.6	32.3	956	17	AG032985	AG032985 Pan trogl
	26	32.6	32.3	1252	12	BG858080	BG858080 1024055F0
	27	32.4	32.1	665	10	BE374590	BE374590 601225377
c	28	32.4	32.1	891	17	AG144954	AG144954 Pan trogl
c	29	32.4	32.1	907	12	BG846266	BG846266 1024012H1
c	30	32.4	32.1	921	17	CNS03TMZ	AL260036 Tetraodon
c	31	32.4	32.1	928	17	AG054567	AG054567 Pan trogl
	32	32.4	32.1	1504	14	BQ644310	BQ644310 AGENCOURT
c	33	32.4	32.1	1580	11	AK018180	AK018180 Mus muscu
c	34	32.2	31.9	803	17	AG042898	AG042898 Pan trogl
	35	32.2	31.9	928	17	CNS01WM2	AL170579 Tetraodon
	36	32.2	31.9	1002	14	BQ674635	BQ674635 AGENCOURT
	37	32	31.7	647	12	BG355986	BG355986 EM1_19_A0
	38	32	31.7	787	14	BQ573025	BQ573025 UI-M-FD0-
	39	32	31.7	829	9	AU124563	AU124563 AU124563
	40	32	31.7	917	17	AG080418	AG080418 Pan trogl
	41	32	31.7	1452	17	AG032979	AG032979 Pan trogl
c	42	31.8	31.5	754	14	BQ804177	BQ804177 WHE3551_C
c	43	31.8	31.5	821	12	BE732483	BE732483 601567671
c	44	31.8	31.5	1718	17	AG064019	AG064019 Pan trogl
	45	31.6	31.3	153	13	BI423588	BI423588 949050A03

ALIGNMENTS

RESULT 1
BE378391
LOCUS BE378391 639 bp mRNA linear EST 21-JUL-2000
DEFINITION 601236718F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3608765 5', mRNA sequence.
ACCESSION BE378391
VERSION BE378391.1 GI:9323856
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 639)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM262 row: c column: 06
High quality sequence stop: 635.
FEATURES Location/Qualifiers
source 1. .639

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3608765"
                     /clone_lib="NIH_MGC_44"
                     /tissue_type="endometrium, adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies)."
BASE COUNT       93 a    220 c    248 g     78 t
ORIGIN

  Query Match             38.4%;  Score 38.8;  DB 10;  Length 639;
  Best Local Similarity   64.4%;  Pred. No. 9.5;
  Matches   58;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      11 GCCGATGGACGACCCACGCCCGCGGTCGCGTGCAGCCGACAGACGGCGCGCCCGGCGCGT 70
           |||   ||||||| |   | |||   | ||||| |||  |||||| || ||| ||||
Db     185 GCCAGAGGACGACTCGGCCTCGCCCCCCGTGCCGCCAGCAGACGCCGAGCCGGGCGACG 244

Qy      71 TGCCGCGCCTCGAGGTGCTGGAGGACCGCG 100
             |  ||| |  | | ||| || |||| |||
Db     245 AGGAGCGGCCGGGGCTGCAGGTGGACTGCG 274


RESULT 2
BF795540/c
LOCUS       BF795540                978 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602259474F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4342603 5',
            mRNA sequence.
ACCESSION   BF795540
VERSION     BF795540.1  GI:12100594
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 978)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt, M.D., Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9958  row: c  column: 20
            High quality sequence stop: 687.
FEATURES             Location/Qualifiers
     source          1. .978
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4342603"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      187 a    303 c    321 g    167 t
ORIGIN

  Query Match             38.2%;  Score 38.6;  DB 12;  Length 978;
  Best Local Similarity   61.4%;  Pred. No. 9.7;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           |||  |   |||  |  || ||||| |||||||||| |||  | | |||||   |||||
Db     975 GAGCGCACCCGCGAACAGAGGACCCGCGCCCGCGGGCGGCCGGACCCGACGCCCGGCGTC 916

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           || | ||||  | | | | | |||   | ||  ||  | | |
Db     915 CCGGCCGCGCCGGCTCCCGTGGAGCACCGGGTGGGGCCCCC 875


RESULT 3
AG030608
LOCUS       AG030608                1189 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-003A18.R, genomic survey sequence.
ACCESSION   AG030608
VERSION     AG030608.1  GI:16557481
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-003A18.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1189)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1189
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-003A18.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       17 a    488 c    540 g     24 t    120 others
ORIGIN

  Query Match             36.0%;  Score 36.4;  DB 17;  Length 1189;
  Best Local Similarity   61.1%;  Pred. No. 29;
  Matches   58;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       6 CGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGG 65
           ||  |||||   | ||  |   |||| ||| || | || |||| | | || |||| |  |
Db     433 CGCGCGCCGCGCGGCGGNCGCGGCCCCCGGGCGCGCGCCGCCGGCGGGCGCCGCGGCGCG 492

Qy      66 CGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCG 100
             |||  | |||| | ||    || || ||  ||||
Db     493 GGCGCCGGCGCGGCCCGGCCGGCCGGCGGCGCGCG 527


RESULT 4
BG809572
```

```
LOCUS       BG809572                888 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  mgct001xal4f Magnaporthe grisea Appressorium Stage cDNA Magnaporthe
            grisea cDNA clone mgct001xal4f 5', mRNA sequence.
ACCESSION   BG809572
VERSION     BG809572.1  GI:14180552
KEYWORDS    EST.
SOURCE      Magnaporthe grisea.
  ORGANISM  Magnaporthe grisea
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Choi,W. and Dean,R.A.
  TITLE     Construction and sequence analysis of an appressorium stage cDNA
            library in the rice blast fungus, Magnaporthe grisea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Ralph A. Dean
            Fungal Genomics Laboratory
            North Carolina State University
            Campus Box 7251, Raleigh, NC 27695, USA
            Tel: 919-513-0020
            Fax: 919-513-0024
            Email: ralph_dean@ncsu.edu
            Seq primer: T3 primer (AATTAACCCTCACTAAAGGG).
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgct001xal4f"
                     /clone_lib="Magnaporthe grisea Appressorium Stage cDNA"
                     /dev_stage="Germinated conidia on apprressorium-inductive
                     surface"
                     /note="Vector: pBlueScript SK(+) Vector; Site_1: EcoRI;
                     Site_2: XhoI; The appressorium formation-specific cDNA
                     library was constructed from conidia germinated for 5-8 hr
                     on an inductive surface. The library contains over 55,000
                     clones with average insert size of 1.5 kbp."
BASE COUNT       22 a    431 c    397 g      4 t     34 others
ORIGIN

  Query Match             35.8%;  Score 36.2;  DB 12;  Length 888;
  Best Local Similarity   60.8%;  Pred. No. 35;
  Matches   59;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       5 CCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCG 64
           |||  | |||  || |   || |||  |||  || | ||  |   | | ||||||||||
Db     623 CCGGCCCCCGCCGGCCCCGCCGCGCGGGCGCGCGCGCGCCCCGCCCGGGCGGCGCGCCCG 682

Qy      65 GCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           ||||  ||| | || ||  | || |  |  || |||
Db     683 CCGCGGCGCCCCCCCGCGCCGCGCCGCCGCCCCCCGC 719


RESULT 5
AG032381/c
LOCUS       AG032381                813 bp    DNA     linear   GSS 01-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-007B17.F, genomic survey sequence.
ACCESSION   AG032381
VERSION     AG032381.1  GI:16559254
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-007B17.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 813)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-007B17.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       33 a    320 c    267 g    164 t     29 others
ORIGIN

  Query Match             34.7%;  Score 35;  DB 17;  Length 813;
  Best Local Similarity   59.6%;  Pred. No. 66;
  Matches   53;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      12 CCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTT 71
           |||   | |   ||   ||||||| |||| ||||||| | |  ||| | || ||| ||
Db     334 CCGGCAGGCCCGCCGNNCCCGCGGGCGGGGGCAGCCGCCGGGGGGCCCCCCNGGCNCGCN 275

Qy      72 GCCGCGCCTCGAGGTGCTGGAGGACCGCG 100
           |  | ||| || ||    || |   || |
Db     274 GGNGGGCCCCGGGGGCGGGGCGCGGCGGG 246


RESULT 6
AG063051/c
LOCUS       AG063051                1020 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-051J10.F, genomic survey sequence.
ACCESSION   AG063051
VERSION     AG063051.1  GI:16614853
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-051J10.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1020)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
```

```
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1020
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-051J10.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       57 a    377 c    385 g     44 t    157 others
ORIGIN

  Query Match             34.7%;  Score 35;  DB 17;  Length 1020;
  Best Local Similarity   60.9%;  Pred. No. 63;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       5 CCGAACGCCGATGGACGACCCACGCCCGC 3GTCGGGTGCAGCCGACAGACGGCGCGCCCG 64
           |||  || ||  | ||| ||| ||||  | | || | || ||   |   |  |||| |||
Db     985 CCGCGCGGCGGCGCACGCCCCGCGCCGCC 2GCCGCGCGCGGCGCGCGCGCCCCGCGNCCG 926

Qy      65 GCGCGTTGCCGCGCCTCGAGGTGCTGGAG 3AC 96
           || ||   | ||||| || |  || || || |
Db     925 GCCCGCGCCGGCGCCGCGCGCGGCGGGTG 3CC 894


RESULT 7
BQ680479
LOCUS       BQ680479                1059 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8209830 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6272780
            5', mRNA sequence.
ACCESSION   BQ680479
VERSION     BQ680479.1  GI:21793158
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1059)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E  Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2451  row: k  column: 21
            High quality sequence stop: 286.
FEATURES             Location/Qualifiers
     source          1. .1059
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6272780"
                     /clone_lib="NIH_MGC_112"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       88 a    194 c    311 g     78 t    388 others
ORIGIN

  Query Match             34.3%;  Score 34.6;  DB 14;  Length 1059;
  Best Local Similarity   64.2%;  Pred. No. 77;
  Matches   52;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           | | ||   |||||  |  || ||   || | ||| |||| ||  ||| | | ||||| |
Db     852 GCGGCCCGCCGCCGCCGCGCGGCCGCGGCCGCCCGGCCGGGCGCGCCCGGCCGGCGGCGGG 911

Qy      61 CCCGGCGCGTTGCCGCGCCTC 81
           ||||| ||   | ||||| |
Db     912 CCCGGGGCCCCGGCGCGCCCC 932


RESULT 8
AV618417
LOCUS       AV618417                260 bp    mRNA    linear   EST 28-NOV-2001
DEFINITION  AV618417 Bos taurus rumen fetus Bos taurus cDNA clone E0RU001G05
            5', mRNA sequence.
ACCESSION   AV618417
VERSION     AV618417.1  GI:9754087
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE     Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
COMMENT     Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E0RU001G05"
                     /clone_lib="Bos taurus rumen fetus"
                     /tissue_type="rumen"
                     /dev_stage="fetus"
                     /lab_host="DH10B"
                     /note="Vector: pZL1; Site_1: SalI; Site_2: NotI"
BASE COUNT       43 a     99 c     86 g     27 t      5 others
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 260;
  Best Local Similarity   67.1%;  Pred. No. 1.1e+02;
  Matches   47;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      24 CCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGA 83
           || ||||  || | |    || ||||||  || ||||||||| | |  |||| | | ||
Db      19 CCCCGCCGCCGNTAGCCGCCAACCGACAACCGCCGCGCCCGGGGAGGAGCCGTGGCCCGN 78

Qy      84 GGTGCTGGAG 93
           || || ||||
Db      79 GGGGCCGGAG 88


RESULT 9
AG082451
LOCUS       AG082451                983 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-079F14.F, genomic survey sequence.
ACCESSION   AG082451
VERSION     AG082451.1  GI:16634253
KEYWORDS    GSS.
```

```
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-079F14.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 983)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-079F14.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      128 a    391 c    439 g     15 t     10 others
ORIGIN

  Query Match             33.7%;  Score 34;  DB 17;  Length 983;
  Best Local Similarity   63.4%;  Pred. No. 1.1e+02;
  Matches   52;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       6 CGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGG 65
           || | | ||| || ||| ||  |||||||| ||   || |||||| | | | | ||  ||
Db     289 CGGAGGGCGAAGGGCGAGCCGGGCCCGCGGCCGCCCGCGGCCGACGGGCCGAGGGCAGGG 348

Qy      66 CGCGTTGCCGCGCCTCGAGGTG 87
            |||  |  | |   ||||| |
Db     349 GGCGGGGGAGGGGGGCGAGGCG 370


RESULT 10
AJ449574
LOCUS       AJ449574                378 bp    mRNA    linear   EST 19-APR-2002
DEFINITION  AJ449574 riken1 Gallus gallus cDNA clone 22h1r1, mRNA sequence.
ACCESSION   AJ449574
VERSION     AJ449574.1  GI:20216795
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   Buerstedde,J.M.
  TITLE     Gallus gallus bursal lymphocyte EST
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES             Location/Qualifiers
     source          1. .378
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /clone="22h1r1"
                     /clone_lib="riken1"
                     /cell_type="bursal lymphocyte"
                     /dev_stage="2-3 weeks old"
                     /note="CB inbred strain"
BASE COUNT       57 a    136 c    140 g     45 t
ORIGIN

  Query Match             33.5%;  Score 33.8;  DB 9;  Length 378;
  Best Local Similarity   62.4%;  Pred. No. 1.4e+02;
  Matches   53;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      17 GGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGC 76
           ||  |||| | |||   |  || | |  || |    ||| |||||  |||||  ||  |
Db      93 GGCTGACCGAGGCCAAAGCCCGCGAGACGCTGCGCAACGCCGCGCTGAGCGCGCAGCTCC 152

Qy      77 GCCTCGAGGTGCTGGAGGACCGCGC 101
           |||  |  |||||| ||| ||||||
Db     153 GCCAGGCCGTGCTGCAGGCCCGCGC 177


RESULT 11
AJ452432
LOCUS       AJ452432                681 bp    mRNA    linear   EST 22-APR-2002
DEFINITION  AJ452432 riken1 Gallus gallus cDNA clone 30o6r1, mRNA sequence.
ACCESSION   AJ452432
VERSION     AJ452432.1  GI:20262528
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 681)
  AUTHORS   Buerstedde,J.M.
  TITLE     Gallus gallus bursal lymphocyte EST
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Buerstedde JM
            Cellular Immunology
            Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES             Location/Qualifiers
     source          1. .681
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /clone="30o6r1"
                     /clone_lib="riken1"
                     /cell_type="bursal lymphocyte"
                     /dev_stage="2-3 weeks old"
                     /note="CB inbred strain"
BASE COUNT      131 a    198 c    262 g     89 t      1 others
ORIGIN

  Query Match             33.5%;  Score 33.8;  DB 9;  Length 681;
  Best Local Similarity   62.4%;  Pred. No. 1.3e+02;
  Matches   53;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      17 GGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGC 76
           ||  |||| | |||   |  || | |  || |    ||| |||||  |||||  ||  |
Db      93 GGCTGACCGAGGCCAAAGCCCGCGAGACGCTGCGCAACGCCGCGCTGAGCGCGCAGCTCC 152

Qy      77 GCCTCGAGGTGCTGGAGGACCGCGC 101
           |||  |  |||||| ||| ||||||
Db     153 GCCAGGCCGTGCTGCAGGCCCGCGC 177


RESULT 12
AQ272425/c
```

```
LOCUS       AQ272425                724 bp    DNA     linear   GSS 03-NOV-1998
DEFINITION  nbxb0027P16f CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0027P16f, DNA sequence
ACCESSION   AQ272425
VERSION     AQ272425.1  GI:3825740
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 136.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="nbxb0027P16f"
                     /clone_lib="CUGI Rice BAC Library"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Rice is one of two most popular grains in the
                     world. Half of the world  population especially those
                     inhabiting highly populated areas of the  humid tropics
                     and subtropics, rely on rice as their primary source of
                     carbohydrate. Monocotyledonous rice is a diploid plant
                     (2n=24) with a  haploid genome equivalent of 431 Mbp
                     (Arumuganathan and Earle, 1991).  The relatively small
                     genome of rice, three times larger than that of
                     Arabidopsis, makes it suitable for genomic studies. In
                     order to  facilitate positional cloning, physical mapping
                     and genome sequencing  of rice, we have constructed a BAC
                     library from Oryza sativa, Nipponbare  variety. The
                     library contains 36,864 clones with an average insert size
                     of 128.5 Kb providing 10.9 haploid genome equivalents. The
                     deep  coverage allows the isolation a particular sequence
                     with a probability  of 99.9 %. Two high density filters,
                     each containing 18,432 clones  (doubly spotted), represent
                     the whole library for colony screening."
BASE COUNT       12 a    287 c    300 g     95 t     30 others
ORIGIN

  Query Match             33.3%;  Score 33.6;  DB 17;  Length 724;
  Best Local Similarity   59.3%;  Pred. No. 1.4e+02;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      11 GCCGATGGACGACCCACGCCCGCGGTCGGCTGCAGCCGACAGACGGCGCGCCCGGCGCGT 70
           |||   || || ||| | | ||| | || | || | | || ||  | ||| |||
Db     418 GCCCCGGGGCGCCCCGCNCGCGCCGCCGCGGCCGGCGGCNGCCGCCGNCCGCGGGGCGC 359

Qy      71 TGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
             || ||| | |  || ||  |  |  || ||
Db     358 GGCGGCGGCCCCGGGCGCCCGCCGGGCGGGC 328


RESULT 13
AG137902
LOCUS       AG137902                813 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-151019.R, genomic survey sequence.
ACCESSION   AG137902
VERSION     AG137902.1  GI:16667580
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-151019.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 813)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .813
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-151019.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      174 a    205 c    303 g     20 t    111 others
ORIGIN

  Query Match             33.3%;  Score 33.6;  DB 17;  Length 813;
  Best Local Similarity   56.4%;  Pred. No. 1.4e+02;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
             | ||| |  | |||  ||  | |||    |   || | |  | || |  | | ||||||
Db     399 GNGACGGCNCNCCGGGGGGGGCCCCGGNNCGCGGGGCCGCCGGAGACCCCCGGCGGCGCC 458

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           |||||| ||  |||   || |     || | | | ||||||
Db     459 CCCGGCCCGGGGCCCGCCCCCCCCCGGCCGCAAGCCCGCGC 499


RESULT 14
CNS006ON
LOCUS       CNS006ON                910 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL065629
VERSION     AL065629.1  GI:4944698
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .910
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="BACR14J21"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
BASE COUNT      202 a     63 c    112 g    198 t    335 others
ORIGIN

  Query Match             33.3%;  Score 33.6;  DB 17;  Length 910;
  Best Local Similarity   28.6%;  Pred. No. 1.3e+02;
  Matches   20;  Conservative   37;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
           |:||| :|: |  :::: | : ::::::::::||:::::| :::||||:|  :: ::|||:||:
Db     538 GYGGTGSGSGGGGSSSSGCSSSSSSSSSSSSSCGSSSBGCBSSCGCGSCSYSSSSTGSTGS 597

Qy      92 AGGACCGCGC 101
             |  ::::::
Db     598 CGCGSSSSSS 607


RESULT 15
BQ788818/c
LOCUS       BQ788818                643 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  WHE4154_B12_D24ZS Wheat CS whole plant cDNA library Triticum
            aestivum cDNA clone WHE4154_B12_D24, mRNA sequence.
ACCESSION   BQ788818
VERSION     BQ788818.1  GI:21997290
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J.
            , Lazo,G.R., Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Chinese Spring whole plant cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES              Location/Qualifiers
     source           1. .643
                      /organism="Triticum aestivum"
                      /cultivar="Chinese Spring"
                      /db_xref="taxon:4565"
                      /clone="WHE4154_B12_D24"
                      /clone_lib="Wheat CS whole plant cDNA library"
                      /tissue_type="Roots, leaves, crown, stem and sheath"
                      /dev_stage="Adult"
                      /lab_host="E. coli SOLR"
                      /note="Vector: Lambda Uni-ZAP XR, excised phagemid
                      pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; Plant
                      tissues from wheat cv. CS grown to full tillering stage in
                      greenhouse were collected at University of California,
                      Davis (Jan Dvorak's lab, K. Deal and E. Akhunov). Total
                      RNA was prepared from leaves (young leaf and third leaf),
                      whole roots, crown, steam and sheath tissues, and then
                      equal quantities of RNA were pooled from the these
                      samples. PolyA was purified from the pooled RNA, a cDNA
                      library was made, and the cDNA clones were in vivo excised
                      to give pBluescript SK(-) phagemids in J. Dvorak's lab (E.
                      Akhunov, J. Dvorak) at the University of California,
                      Davis. Colony plating, plasmid DNA preparations and DNA
                      sequencing were performed in the OD Anderson lab (all
                      other authors)."
BASE COUNT      120 a    211 c    218 g     94 t
ORIGIN

  Query Match             32.9%;  Score 33.2;  DB 14;  Length 643;
  Best Local Similarity   61.6%;  Pred. No. 1.8e+02;
  Matches   53;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       5 CCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCG 64
             ||| ||  ||  |||||     |||| |  | ||| ||| || ||| | ||||   | ||
Db     207 CCGCACCGCGGAGGACGCGGAGCGCCGGGCGACGGATGCCGCAGACCGCCGGCCGCCTCG 148

Qy      65 GCGCGTTGCCGCGCCTCGAGGTGCTG 90
           ||||   ||  | || | | |||||
Db     147 CCGCGAGCCCCTGTCTGGCGATGCTG 122


Search completed: November  5, 2002, 16:18:18
Job time : 1118.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_28200_28300
Perfect score: 101
Sequence: 1 gagaccgaacgccgatggac.....gaggtgctggaggaccgcgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	101	100.0	71989	21	AAA29349	Sorangium cellulos
c	3	32.6	32.3	384	24	ABN25960	Human ORFX polynuc
c	4	32.2	31.9	88421	24	AAL40781	88421nt genomic DN
	5	32	31.7	829	22	AAH05398	Human cDNA clone (
	6	32	31.7	2210	22	AAH15833	Human cDNA sequenc
	7	31.2	30.9	1128	20	AAZ10560	DNA encoding G pro
	8	31.2	30.9	1128	21	AAD01125	Human orphan G pro
	9	31.2	30.9	1128	21	AAA46026	Human G protein co
	10	31.2	30.9	1128	24	AAD27892	Human G-protein co
	11	31.2	30.9	1128	24	AAS98049	Human DNA for pote
c	12	31	30.7	2573	19	AAV46297	C. reinhardtii pro
c	13	31	30.7	3381	19	AAV46298	C. reinhardtii pro
	14	30.8	30.5	465	24	ABN17674	Human ORFX polynuc
c	15	30.8	30.5	1899	24	ABQ90027	M. capsulatus gene
c	16	30.4	30.1	881	24	ABQ29954	Oligonucleotide fo
	17	30.4	30.1	881	24	ABQ29955	Oligonucleotide fo
c	18	30	29.7	2520	24	ABK84011	Human cDNA differe
c	19	30	29.7	2520	24	ABL61822	Colon adenocarcino
	20	29.8	29.5	1797	22	AAA54514	Fructan exohydrola
	21	29.8	29.5	33529	23	AAS17367	DNA sequence of S.
c	22	29.6	29.3	1500	22	AAD19383	Mouse CCAAT/enhanc
c	23	29.6	29.3	1500	23	ABK51405	DNA encoding mouse
c	24	29.6	29.3	1500	24	AAL44080	Mouse C/EBP-beta p
c	25	29.6	29.3	1500	24	ABK51994	DNA encoding wild-
c	26	29.6	29.3	1500	24	ABL94244	Mouse C/EBP beta-e
	27	29.6	29.3	4430	23	AAS86303	DNA encoding novel
	28	29.6	29.3	53789	19	AAV21187	Amycolatopsis medi
c	29	29.4	29.1	1350	23	AAS69342	DNA encoding novel
	30	29.4	29.1	1596	13	AAQ21975	DNA encoding HSV-1
c	31	29.4	29.1	10211	19	AAV62152	HSV-2 strain SB5 C
c	32	29.4	29.1	117213	19	AAV62176	HSV-2 strain SB5 C
c	33	29.4	29.1	154746	24	AAD25519	Human herpesvirus
c	34	29.2	28.9	2053	14	AAQ46816	DTA gene. Xanthom
	35	29.2	28.9	5609	24	ABN59769	Novel human coding
	36	29	28.7	287	22	AAI68354	Rice s11-2 signal
c	37	29	28.7	5801	22	ABA18280	Human nervous syst
	38	28.8	28.5	300	20	AAZ13774	Human gene express
	39	28.8	28.5	4566	21	AAC74710	Human ORFX ORF265
	40	28.8	28.5	4607	23	ABV28262	Human prostate exp
c	41	28.8	28.5	8169	19	AAV26609	Actinomadura hibis
c	42	28.6	28.3	656	21	AAC77728	Human cancer assoc
c	43	28.6	28.3	856	24	ABQ24000	Oligonucleotide fo
	44	28.6	28.3	856	24	ABQ24001	Oligonucleotide fo
	45	28.6	28.3	897	18	AAT79335	DNA encoding Melit

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;    99WO-EP04171.
XX
PR   18-JUN-1998;    98US-0099504.
PR   24-SEP-1998;    98US-0101631.
PR   05-FEB-1999;    99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   33812 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 33871

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   33872 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 33912


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes AEP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes K36"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AP6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KP6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes AEP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 60
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   28200 GAGACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCG 28259

Qy      61 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   28260 CCCGGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCGCGC 28300


RESULT 3
ABN25960/c
ID   ABN25960 standard; cDNA; 384 BP.
XX
AC   ABN25960;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:20397.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP10208.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 20397; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 384 BP; 75 A; 135 C; 98 G; 74 T; 2 other;

  Query Match             32.3%;  Score 32.6;  DB 24;  Length 384;
  Best Local Similarity   58.9%;  Pred. No. 14;
  Matches   56;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       4 ACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCC 63
             | ||| ||| | | |  |||| | || | |||  | |||| |||  | |  |  |   |
Db     262 AACGAGCGCGGCTCGGTGACCGATGCACCCGGCGGCGTGCTGCCTTCGGGGGTGGACGCG 203

Qy      64 GGCGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCG 98
           |||||   | |  ||| ||||||||| ||  ||
Db     202 ATGGCGTTAGTGGGGGTCGTGGTGCTGGGGGTTCG 168


RESULT 4
AAL40781/c
ID   AAL40781 standard; DNA; 88421 BP.
XX
AC   AAL40781;
XX
DT   03-OCT-2002  (first entry)
XX
DE   88421nt genomic DNA of ramoplanin producing Actinoplanes sp.
```

```
XX
KW   Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;
KW   biosynthesis gene cluster; bioengineering; peptide synthetase module;
KW   adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;
KW   chlorinate; lipdepsipeptide; gene; ds.
XX
OS   Actinoplanes sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             2077..3078
FT                   /*tag= a
FT                   /product= "Protein of ORF 1"
FT   CDS             3118..4032
FT                   /*tag= b
FT                   /product= "Protein of ORF 2"
FT   CDS             4038..5048
FT                   /*tag= c
FT                   /product= "Protein of ORF 3"
FT   CDS             complement (6665..5814)
FT                   /*tag= d
FT                   /product= "Protein of ORF 4"
FT   CDS             complement (7703..6693)
FT                   /*tag= e
FT                   /product= "Protein of ORF 5"
FT   CDS             complement (9464..8130)
FT                   /*tag= f
FT                   /product= "Protein of ORF 6"
FT   CDS             9691..10761
FT                   /*tag= g
FT                   /product= "Protein of ORF 7"
FT   CDS             complement (12751..10829)
FT                   /*tag= h
FT                   /product= "Protein of ORF 8"
FT   CDS             complement (13617..12802)
FT                   /*tag= i
FT                   /product= "Protein of ORF 9"
FT   CDS             complement (15203..13614)
FT                   /*tag= j
FT                   /product= "Protein of ORF 10"
FT   CDS             15591..15863
FT                   /*tag= k
FT                   /product= "Protein of ORF 11"
FT   CDS             15880..19035
FT                   /*tag= l
FT                   /product= "Protein of ORF 12"
FT   CDS             19032..39713
FT                   /*tag= m
FT                   /product= "Protein of ORF 13"
FT   CDS             39713..65800
FT                   /*tag= n
FT                   /product= "Protein of ORF 14"
FT   CDS             65826..66530
FT                   /*tag= o
FT                   /product= "Protein of ORF 15"
FT   CDS             66546..67370
FT                   /*tag= p
FT                   /product= "Protein of ORF 16"
FT   CDS             67384..70059
FT                   /*tag= q
FT                   /product= "Protein of ORF 17"
FT   CDS             70099..70662
FT                   /*tag= r
FT                   /product= "Protein of ORF 18"
FT   CDS             70659..71906
FT                   /*tag= s
FT                   /product= "Protein of ORF 19"
FT   CDS             complement (73439..71964)
FT                   /*tag= t
FT                   /product= "Protein of ORF 20"
FT   CDS             complement (74216..73563)
FT                   /*tag= u
FT                   /product= "Protein of ORF 21"
FT   CDS             complement (75424..74213)
FT                   /*tag= v
FT                   /product= "Protein of ORF 22"
FT   CDS             75535..76464
FT                   /*tag= w
FT                   /product= "Protein of ORF 23"
FT   CDS             complement (78110..76449)
FT                   /*tag= x
FT                   /product= "Protein of ORF 24"
FT   CDS             complement (79864..78107)
FT                   /*tag= y
FT                   /product= "Protein of ORF 25"
FT   CDS             complement (81624..79861)
FT                   /*tag= z
FT                   /product= "Protein of ORF 26"
FT   CDS             complement (81909..81682)
FT                   /*tag= aa
FT                   /product= "Protein of ORF 27"
FT   CDS             complement (82346..82062)
FT                   /*tag= ab
FT                   /product= "Protein of ORF 28"
FT   CDS             82587..84446
FT                   /*tag= ac
FT                   /product= "Protein of ORF 29"
FT   CDS             84481..85548
FT                   /*tag= ad
FT                   /product= "Protein of ORF 30"
FT   CDS             85556..86845
FT                   /*tag= ae
FT                   /product= "Protein of ORF 31"
FT   CDS             87372..86803
FT                   /*tag= af
FT                   /product= "Protein of ORF 32"
FT   CDS             87494..88420
FT                   /*tag= ag
FT                   /product= "Protein of ORF 33"
XX
PN   WO200231155-A2.
XX
PD   18-APR-2002.
XX
PF   15-OCT-2001; 2001WO-CA01462.
XX
PR   13-OCT-2000; 2000US-239924P.
PR   12-APR-2001; 2001US-283296P.
PR   24-JUL-2001; 2001US-0910813.
XX
PA   (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI   Farnet CM,  Zazopoulos E,  Staffa A;
XX
DR   WPI; 2002-435445/46.
DR   P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,
DR   AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,
DR   AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,
DR   AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,
DR   AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.
XX
PT   Novel isolated ramoplanin biosynthetic pathway polypeptide useful for
PT   chemically modifying biological molecule that is a substrate for a \
PT   polypeptide encoded by a ramoplanin biosynthesis gene cluster -
XX
PS   Disclosure; Page 87-135; 212pp; English.
XX
CC   The invention relates to an isolated ramoplanin biosynthetic pathway
CC   polypeptide selected from a polypeptide of open reading frames (ORF)
CC   1-32. The isolated polypeptides are useful for chemically modifying a
CC   biological molecule that is a substrate for a polypeptide encoded by a
CC   ramoplanin biosynthesis gene cluster, by contacting the biological
CC   molecule with the isolated polypeptide, where the polypeptide chemically
CC   modifies the biological molecule. The method comprises contacting the
CC   biological molecule with at least two different polypeptides encoded by
CC   ramoplanin ORFs 1-31. The polypeptides are useful for directing the
CC   biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated
```

CC gene cluster comprising the ORFs is useful as a substrate for
CC bioengineering of antibiotic structures. An isolated polypeptide or its
CC encoding nucleic acid sequence is useful for generating derivatives of
CC ramoplanin, for improving production or for producing variants of other
CC antibiotics of the peptide class. The isolated polypeptides are useful
CC for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain
CC in conjunction with other peptide synthetase modules and allowing the
CC incorporation of Thr into a peptide antibiotic precursor, for modifying
CC fatty acid structure and/or enhancing fatty acid incorporation into the
CC peptide antibiotic structure, for production of an hydroxyphenylglycine
CC (HPG)-containing peptide antibiotic, for enhancing secretion of
CC ramoplanin or its variants and derivatives, for enhancing uptake of
CC precursors for ramoplanin biosynthesis, for enhancing production of
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of
CC a peptide antibiotic precursor, and for designing specific nucleotide
CC probes and primers for identifying and isolating putative lipdepsipeptide
CC -producing microorganisms. This polynucleotide sequence represents the
CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.
CC microorganism of the invention.
XX
SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 other;

Query Match 31.9%; Score 32.2; DB 24; Length 88421;
Best Local Similarity 59.1%; Pred. No. 13;
Matches 55; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 6 CGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGG 65
Db 78265 CGGGCGCCGGCGCCCGATCCCGGCCCGCACGACGACGGCGCCGACCGGCTGCTGGCCGCG 78206

Qy 66 CGCGTTGCCGCGCCTCGAGGTGCTGGAGGACCG 98
Db 78205 CGGGTCGCGGCGGCCGCACGCACCGGTGGACCG 78173

RESULT 5
AAH05398
ID AAH05398 standard; cDNA; 829 BP.
XX
AC AAH05398;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA clone (5'-primer) SEQ ID NO:2233.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 1; SEQ ID 2233; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 829 BP; 145 A; 292 C; 239 G; 149 T; 4 other;

Query Match 31.7%; Score 32; DB 22; Length 829;
Best Local Similarity 60.2%; Pred. No. 19;
Matches 53; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 12 CCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTT 71
Db 94 CCGAGTGGCCTACCGAGCCCGAGTGGCCCCGCAGCGTCCAGGAGGCGCCCGCTCCGCGGT 153

Qy 72 GCCGCGCCTCGAGGTGCTGGAGGACCGC 99
Db 154 GGCGCTCTTGGAGGTGGTGTCGGAGAGC 181

RESULT 6
AAH15833
ID AAH15833 standard; cDNA; 2210 BP.
XX
AC AAH15833;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14325.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX

```
PT   Primer sets for synthesizing polynucleotides, particularly the 5602
PT   full-length cDNAs defined in the specification, and for the detection
PT   and/or diagnosis of the abnormality of the proteins encoded by the
PT   full-length cDNAs -
XX
PS   Claim 8; SEQ ID 14325; 2537pp + CD ROM; English.
XX
CC   The present invention describes primer sets for synthesising 5602
CC   full-length cDNAs defined in the specification. Where a primer set
CC   comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC   to the complementary strand of a polynucleotide which comprises one of
CC   the 5602 nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in
CC   the specification. The primer sets can be used in antisense therapy and
CC   in gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC   AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC   represent oligonucleotides, all of which are used in the exemplification
CC   of the present invention.
XX
SQ   Sequence 2210 BP; 503 A; 643 C; 592 G; 472 T; 0 other;

  Query Match             31.7%;  Score 32;  DB 22;  Length 2210;
  Best Local Similarity   60.2%;  Pred. No. 18;
  Matches   53;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      12 CCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTT 71
           ||||  | |   ||  ||||| |       |||||   |||  ||||| | |  |||| |
Db      94 CCGAGTGGCCTACCGAGCCCGAGTGGCCCCGCAGCGTCCAGGAGGCGCCCGCTCCGCGGT 153

Qy      72 GCCGCGCCTCGAGGTGCTGGAGGACCGC 99
           | ||| | | |||||| ||  |||  ||
Db     154 GGCGCTCTTGGAGGTGGTGTCGGAGAGC 181


RESULT 7
AAZ10560
ID   AAZ10560 standard; DNA; 1128 BP.
XX
AC   AAZ10560;
XX
DT   15-NOV-1999  (first entry)
XX
DE   DNA encoding G protein-coupled receptor protein designated SREB1.
XX
KW   G protein-coupled receptor protein; SREB1; central nervous system;
KW   inflammatory disorder; immunological; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1128
FT                   /*tag=  a
FT                   /product= "SREB1"
XX
PN   WO9946378-A1.
XX
PD   16-SEP-1999.
XX
PF   11-MAR-1999;   99WO-JP01191.
XX
PR   12-MAR-1998;   98JP-0060245.
PR   03-FEB-1999;   99JP-0026774.
XX
PA   (YAMA ) YAMANOUCHI PHARM CO LTD.
XX
PI   Matsumoto M,  Sugimoto T,  Takasaki J,  Saito T, 'Kamohara M;
XX
DR   WPI; 1999-551407/46.
DR   P-PSDB; AAY30532.
XX
PT   G protein-coupled receptor proteins expressed in the central nervous
PT   system and genes encoding them  -
XX
PS   Example 1; Page 47-49; 72pp; Japanese.
XX
CC   The present sequence encodes a G protein-coupled receptor protein,
CC   designated SREB1. The protein is expressed in the central nervous
CC   system. The SREB products are used for the diagnosis and treatment
CC   of diseases of the central nervous system, including inflammatory
CC   disorders of immunological origin.
XX
SQ   Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;

  Query Match             30.9%;  Score 31.2;  DB 20;  Length 1128;
  Best Local Similarity   70.0%;  Pred. No. 29;
  Matches   42;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
           |||  | || ||||| | | |||||||| ||||||||| ||     |||   | ||||||
Db     512 GCGCCCTGGAGCAGCGGCCCGACGGCGCCCCCGGCGCGCTGGGCTTCCTGCTGCTGCTGG 571


RESULT 8
AAD01125
ID   AAD01125 standard; cDNA; 1128 BP.
XX
AC   AAD01125;
XX
DT   02-NOV-2000  (first entry)
XX
DE   Human orphan G protein-coupled receptor hGPCR27 cDNA.
XX
KW   Human; orphan G protein-coupled receptor; GPCR; hGPCR27; drug screening;
KW   transmembrane receptor; expressed sequence tag; EST; signal cascade; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1128
FT                   /*tag=  a
FT                   /product= "hGPCR27"
FT                   /note= "Human orphan G protein-coupled receptor"
XX
PN   WO200031258-A2.
XX
PD   02-JUN-2000.
XX
PF   13-OCT-1999;   99WO-US23687.
XX
PR   20-NOV-1998;   98US-0109213.
PR   16-FEB-1999;   99US-0120416.
PR   26-FEB-1999;   99US-0121852.
PR   12-MAR-1999;   99US-0123946.
PR   12-MAR-1999;   99US-0123949.
PR   28-MAY-1999;   99US-0136436.
PR   28-MAY-1999;   99US-0136437.
PR   28-MAY-1999;   99US-0136439.
PR   28-MAY-1999;   99US-0136567.
PR   28-MAY-1999;   99US-0137127.
PR   28-MAY-1999;   99US-0137131.
PR   29-JUN-1999;   99US-0141448.
PR   29-SEP-1999;   99US-0156555.
PR   29-SEP-1999;   99US-0156633.
PR   29-SEP-1999;   99US-0156634.
PR   29-SEP-1999;   99US-0156653.
```

PR 01-OCT-1999; 99US-0157280.
PR 01-OCT-1999; 99US-0157281.
PR 01-OCT-1999; 99US-0157282.
PR 01-OCT-1999; 99US-0157293.
PR 01-OCT-1999; 99US-0157294.
PR 12-OCT-1999; 99US-0416760.
PR 12-OCT-1999; 99US-0417044.
XX
PA (AREN-) ARENA PHARM INC.
XX
PI Chen R, Dang HT, Liaw CW, Lin I;
XX
DR WPI; 2000-400068/34.
DR P-PSDB; AAY71298.
XX
PT Novel human orphan G protein-coupled receptors and the encoding cDNAs
PT for use in the identification of G protein-coupled receptor agonists -
XX
PS Claim 29; Page 61-62; 102pp; English.
XX
CC The present sequence is a cDNA encoding hGPCR27, an endogenous human
CC orphan G protein-coupled receptor (GPCR), expressed in foetal brain,
CC putamen, pituitary gland and caudate nucleus. The hGPCR27 cDNA was
CC identified using EST (expressed sequence tag) AA775870 as a probe.
CC The orphan GPCR of the invention, like all GPCRs has seven transmembrane
CC alpha helices with an extracellular N-terminus and an intracellular
CC C-terminus. However, no endogenous ligands has yet been identified for
CC the proteins of the invention. The orphan GPCRs may be used in the
CC identification of their endogenous ligands, and to screen potential GPCR
CC agonists and antagonists for use as pharmaceutical agents. The proteins
CC may also be used in the study of GPCR-mediated signalling cascades, and
CC to elucidate their precise role in normal and diseased human conditions.
CC Nucleic acid encoding human orphan GPCRs may be used for tissue
CC localisation expression analysis to provide information about their
CC function in healthy and pathological states.
XX
SQ Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;

Query Match 30.9%; Score 31.2; DB 21; Length 1128;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
Db 512 GCGCCCTGGAGCAGCGGCCCGACGGCGCCCCGGCGCGCTGGGCTTCCTGCTGCTGCTGG 571

RESULT 9
AAA46026
ID AAA46026 standard; cDNA; 1128 BP.
XX
AC AAA46026;
XX
DT 22-AUG-2000 (first entry)
XX
DE Human G protein coupled receptor hGPCR27 encoding cDNA SEQ ID NO:17.
XX
KW Human; G protein coupled receptor; GPCR; transmembrane receptor;
KW identification; agonist; screening; therapeutic; pharmaceutical;
KW mutant; ss.
XX
OS Homo sapiens.
XX
PN WO200022131-A2.
XX
PD 20-APR-2000.
XX
PF 13-OCT-1999; 99WO-US24065.
XX
PR 13-OCT-1998; 98US-0170496.
PR 12-NOV-1998; 98US-0108029.
PR 20-NOV-1998; 98US-0109213.
PR 27-NOV-1998; 98US-0110060.
PR 16-FEB-1999; 99US-0120416.
PR 26-FEB-1999; 99US-0121852.
PR 12-MAR-1999; 99US-0123944.
PR 12-MAR-1999; 99US-0123945.
PR 12-MAR-1999; 99US-0123946.
PR 12-MAR-1999; 99US-0123948.
PR 12-MAR-1999; 99US-0123949.
PR 12-MAR-1999; 99US-0123951.
PR 28-MAY-1999; 99US-0136436.
PR 28-MAY-1999; 99US-0136437.
PR 28-MAY-1999; 99US-0136439.
PR 28-MAY-1999; 99US-0137127.
PR 28-MAY-1999; 99US-0137131.
PR 28-MAY-1999; 99US-0137567.
PR 30-JUN-1999; 99US-0141448.
PR 27-AUG-1999; 99US-0151114.
PR 03-SEP-1999; 99US-0152524.
PR 29-SEP-1999; 99US-0156633.
PR 29-SEP-1999; 99US-0156555.
PR 29-SEP-1999; 99US-0156634.
XX
PA (AREN-) ARENA PHARM INC.
XX
PI Behan DP, Lehmann-Bruinsma K, Chalmers DT, Chen R, Dang HT;
PI Gore M, Liaw CW, Lin I, Lowitz K, White C;
XX
DR WPI; 2000-317986/27.
DR P-PSDB; AAB02832.
XX
PT Non-endogenous, human G protein-coupled receptors for screening
PT receptor, inverse or partial agonists useful as therapeutic agents -
XX
PS Example 1; Page 91-92; 187pp; English.
XX
CC The present invention describes transmembrane receptors, preferably
CC human G protein coupled receptors (GPCR), for which the endogenous
CC ligand is unknown (orphan GPCR receptors). More specifically the present
CC invention relates to non-endogenous, constitutively activated versions
CC of a human GPCR. These non-endogenous human GPCRs can be useful for
CC the direct identification of candidate compounds as receptors agonists,
CC inverse agonists or partial agonists for use as pharmaceutical agents.
CC AAA46017 to AAA46126 and AAB02825 to AAB02859 represent sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;

Query Match 30.9%; Score 31.2; DB 21; Length 1128;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 42; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

Qy 32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
Db 512 GCGCCCTGGAGCAGCGGCCCGACGGCGCCCCCGGCGCGCTGGGCTTCCTGCTGCTGCTGG 571

RESULT 10
AAD27892
ID AAD27892 standard; DNA; 1128 BP.
XX
AC AAD27892;
XX
DT 31-MAY-2002 (first entry)
XX
DE Human G-protein coupled receptor 14266 DNA.
XX
KW Human; G-protein coupled receptor; GPCR 14266; haematopoietic disorder;
KW neutrophil deficiency disorder; splenomegaly; pulmonary hypertension;
KW colon; diarrhoea; hepatic injury; idiopathic inflammatory bowel disease;
KW uterus; endometriosis; brain; acute meningitis; multiple sclerosis;
KW T-cell; systemic lupus erythematosus; skin; vitiligo; heart failure;
KW angina pectoris; atherosclerosis; haemolytic anaemia; thymic cyst; ds.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1128
FT                   /*tag= a
FT                   /product= "GPCR 14266"
FT   sig_peptide     1..102
FT                   /*tag= b
FT   mat_peptide     103..1125
FT                   /*tag= c
FT                   /product= "Mature GPCR 14266"
XX
PN   WO200212344-A2.
XX
PD   14-FEB-2002.
XX
PF   08-AUG-2001; 2001WO-US24835.
XX
PR   09-AUG-2000; 2000US-0634392.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Weich NS;
XX
DR   WPI; 2002-217181/27.
DR   P-PSDB; AAE14597.
XX
PT   Identifying compound that binds to and/or modulate activity of the
PT   polypeptide for treating splenomegaly, emphysema and multiple
PT   sclerosis, comprises using human G-protein coupled receptor 14266
PT   polypeptide -
XX
PS   Claim 1; Fig 2; 115pp; English.
XX
CC   The invention relates to a method of identifying an agent that binds to
CC   and/or modulates activity of human G-protein coupled receptor (GPCR)
CC   14266. The invention also provides a method for modulating the level
CC   or activity of the GPCR in cells in vivo or in vitro, or in a subject
CC   predisposed to having a haematopoietic or neutrophil deficiency
CC   disorders. Detecting 14266 polypeptides or nucleic acids in specific
CC   biological samples is useful in disease diagnosis. Modulating level or
CC   activity of the GPCR polypeptide or polynucleotide is useful for treating
CC   diseases related to 14266 receptor malfunction e.g., disorders involving
CC   spleen (e.g. splenomegaly, neoplasms); lung (e.g. pulmonary hypertension,
CC   emphysema); colon (e.g. diarrhoea and dysentery, idiopathic inflammatory
CC   bowel disease); liver (e.g. hepatic injury, viral hepatitis); uterus and
CC   endometrium (e.g. endometriosis, endometrial polyps); brain (e.g.
CC   acute meningitis, multiple sclerosis, spinocerebellar degeneration);
CC   T-cells (e.g. systemic lupus erythematosus, polyarteritis nodosa); skin
CC   (e.g. vitiligo, acanthosis nigricans); heart (e.g. heart failure,
CC   angina pectoris); blood vessels (e.g. atherosclerosis, Raynaud disease,
CC   thalassaemia syndromes, haemolytic anaemia) and disorders of the thymus
CC   (e.g. thymic cysts, Hodgkin's disease). The present sequence is human
CC   GPCR 14266 DNA.
XX
SQ   Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;

  Query Match             30.9%;  Score 31.2;  DB 24;  Length 1128;
  Best Local Similarity   70.0%;  Pred. No. 29;
  Matches   42;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
           |||  | || ||||| | | |||||||| ||||||||| ||     |||   | ||||||
Db     512 GCGCCCTGGAGCAGCGGCCCGACGGCGCCCCCGGCGCGCTGGGCTTCCTGCTGCTGCTGG 571


RESULT 11
AAS98049
ID   AAS98049 standard; DNA; 1128 BP.
XX
AC   AAS98049;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Human DNA for potential G protein-coupled receptor #7.
XX
KW   Human; G protein-coupled receptor; GPCR; ds; GAL4; galanin receptor;
KW   Alzheimer's disease; amyotrophic lateral sclerosis; asthma;
KW   atherosclerosis; basal cell carcinoma; breast carcinoma; cardiomyopathy;
KW   chondrosarcoma; chronic obstructive pulmonary disease; Crohn's disease;
KW   depression; epilepsy; macular degeneration; lymphoma; melanoma;
KW   multiple sclerosis; osteoarthritis; osteoporosis; Parkinson's disease;
KW   psoriasis; rheumatoid arthritis; schizophrenia; ulcerative colitis;
KW   tuberculosis; cognition disorder; memory disorder; anorexia;
KW   hormonal release disorder; cardiovascular activity disorder;
KW   pain perception disorder; obesity; diabetes; obesity;
KW   diabetes; hyperlipidaemia; stroke; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200185791-A1.
XX
PD   15-NOV-2001.
XX
PF   11-MAY-2001; 2001WO-US15332.
XX
PR   11-MAY-2000; 2000US-203217P.
PR   18-MAY-2000; 2000US-205945P.
XX
PA   (LIFE-) LIFESPAN BIOSCIENCES INC.
XX
PI   Brown JP,  Miller M,  Burmer G,  Fabre-Suver C,  Pritchard D;
XX
DR   WPI; 2002-066595/09.
XX
PT   Novel G protein-coupled receptor polypeptides including galanin
PT   receptor polypeptides useful for identifying modulators that are useful
PT   for treating Alzheimer's disease, psoriasis, melanoma, multiple
PT   sclerosis, stroke -
XX
PS   Disclosure; Page 116-117; 144pp; English.
XX
CC   The invention relates to an isolated polypeptide encoded by a
CC   nucleic acid molecule that is at least 80% identical to the G
CC   protein-coupled (GPCR) polynucleotides included in the specification.
CC   Also included are probes based on the GPCR sequences (including
CC   antisense probes), a host cell comprising an expression vector comprising
CC   the GPCR sequence, antibodies raised against the polypeptides,
CC   and methods of identifying modulators of the polypeptides. The
CC   polypeptides are useful for identifying modulator compounds which
CC   function as modulators, activators, repressors, agonists or antagonists
CC   of the novel GPCR polypeptides including the GAL4 polypeptide. The
CC   antibodies and nucleic acid probes as described above can be used to
CC   detect the presence of the polypeptides and nucleic acids and are used to
CC   diagnose a variety of diseases or disorders in which GPCRs are involved
CC   e.g., Alzheimer's disease, amyotrophic lateral sclerosis, asthma,
CC   atherosclerosis, basal cell carcinoma, breast carcinoma, cardiomyopathy,
CC   chondrosarcoma, chronic obstructive pulmonary disease, Crohn's disease,
CC   depression, epilepsy, macular degeneration, lymphoma, melanoma,
CC   multiple sclerosis, osteoarthritis, osteoporosis, Parkinson's disease,
CC   psoriasis, rheumatoid arthritis, schizophrenia, ulcerative colitis,
CC   tuberculosis and many other diseases listed in the specification. The
CC   probes and antibodies are also useful for diagnosing cognition and memory
CC   disorders, anorexia, hormonal release disorders, cardiovascular activity
CC   disorders, pain perception disorders, obesity, diabetes, Alzheimer's
CC   disease. Preferably, compounds that decrease or increase
CC   the expression of galanin receptor (GAL4) can be used to treat obesity,
CC   diabetes, hyperlipidaemia and stroke. The GPCR nucleic acid is
CC   useful for treating the  above mentioned disorders by gene therapy
CC   techniques. The present sequence is a novel GPCR polynucleotide of the
CC   invention.
XX
SQ   Sequence 1128 BP; 116 A; 437 C; 395 G; 180 T; 0 other;

  Query Match             30.9%;  Score 31.2;  DB 24;  Length 1128;
  Best Local Similarity   70.0%;  Pred. No. 29;
  Matches   42;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;
```

```
Qy      32 GCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCGAGGTGCTGG 91
           ||| | || ||||| | | |||||||| |||||||| ||     |||   | ||||||
Db     512 GCGCCCTGGAGCAGCGGCCCGACGGCGCCCCCGGCGCGCTGGGCTTCCTGCTGCTGCTGG 571


RESULT 12
AAV46297/c
ID   AAV46297 standard; DNA; 2573 BP.
XX
AC   AAV46297;
XX
DT   16-OCT-1998  (first entry)
XX
DE   C. reinhardtii protoporphyrinogen oxidase DNA fragment Xho/Pmac2.6.
XX
KW   Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;
KW   crop plant; selective weed control; ss.
XX
OS   Chlamydomonas reinhardtii.
XX
PN   WO9829554-A1.
XX
PD   09-JUL-1998.
XX
PF   27-DEC-1996;   96WO-US20415.
XX
PR   27-DEC-1996;   96WO-US20415.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
PA   (UYDU-) UNIV DUKE.
XX
PI   Boynton JE,  Gillham NW,  Ishige F,  Randolph-Anderson BL;
PI   Sato R;
XX
DR   WPI; 1998-388134/33.
XX
PT   Conferring resistance to herbicides that inhibit proto-porphyrinogen
PT   oxidase - by introducing DNA that encodes a resistant form of the
PT   enzyme, also resistant plants and algae, mutant DNA, vectors and
PT   transformed microorganisms
XX
PS   Example 7; Page 70-71; 108pp; English.
XX
CC   This genomic DNA sequence encodes protoporphyrinogen oxidase from
CC   Chlamydomonas reinhardtii (fragment Xho/Pmac2.6). This sequence is used
CC   in a method which investigates resistance to herbicides that inhibit
CC   protoporphyrinogen oxidase (PPO) using a porphyrin herbicide resistance
CC   domain. The method is applied to crop plants, allowing use of herbicides
CC   for selective weed control. The method can be used to identify compounds
CC   that are, or are not, PPO-inhibitors.
XX
SQ   Sequence 2573 BP; 470 A; 859 C; 735 G; 509 T; 0 other;

  Query Match             30.7%;  Score 31;  DB 19;  Length 2573;
  Best Local Similarity   62.0%;  Pred. No. 31;
  Matches   49;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      23 CCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCG 82
           |||||   || |  |||||||||| ||| || | |  || |||  |  | |   |||| |
Db     400 CCCACCTGCGTGCGCGGGTGCAGCTGACCGAAGCCCGGCACGGACCCGTCCGAGGCCTTG 341

Qy      83 AGGTGCTGGAGGACCGCGC 101
           | | ||   | || ||||
Db     340 CGCTCCTCCCGCACGGCGC 322


RESULT 13
AAV46298/c
ID   AAV46298 standard; DNA; 3381 BP.
XX
AC   AAV46298;
XX
DT   16-OCT-1998  (first entry)
XX
DE   C. reinhardtii protoporphyrinogen oxidase DNA fragment Xho3.4.
XX
KW   Herbicide resistance; porphyrin; protoporphyrinogen oxidase; PPO;
KW   crop plant; selective weed control; ss.
XX
OS   Chlamydomonas reinhardtii.
XX
PN   WO9829554-A1.
XX
PD   09-JUL-1998.
XX
PF   27-DEC-1996;   96WO-US20415.
XX
PR   27-DEC-1996;   96WO-US20415.
XX
PA   (SUMO ) SUMITOMO CHEM CO LTD.
PA   (UYDU-) UNIV DUKE.
XX
PI   Boynton JE,  Gillham NW,  Ishige F,  Randolph-Anderson BL;
PI   Sato R;
XX
DR   WPI; 1998-388134/33.
XX
PT   Conferring resistance to herbicides that inhibit proto-porphyrinogen
PT   oxidase - by introducing DNA that encodes a resistant form of the
PT   enzyme, also resistant plants and algae, mutant DNA, vectors and
PT   transformed microorganisms
XX
PS   Example 7; Page 83-85; 108pp; English.
XX
CC   This genomic DNA sequence encodes protoporphyrinogen oxidase from
CC   Chlamydomonas reinhardtii (fragment Xho3.4). This sequence is used
CC   in a method which investigates resistance to herbicides that inhibit
CC   protoporphyrinogen oxidase (PPO) using a porphyrin herbicide resistance
CC   domain. The method is applied to crop plants, allowing use of herbicides
CC   for selective weed control. The method can be used to identify compounds
CC   that are, or are not, PPO-inhibitors.
XX
SQ   Sequence 3381 BP; 644 A; 1058 C; 1010 G; 669 T; 0 other;

  Query Match             30.7%;  Score 31;  DB 19;  Length 3381;
  Best Local Similarity   62.0%;  Pred. No. 31;
  Matches   49;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      23 CCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCCGGCGCGTTGCCGCGCCTCG 82
           |||||   || |  |||||||||| ||| || | |  || |||  |  | |   |||| |
Db     400 CCCACCTGCGTGCGCGGGTGCAGCTGACCGAAGCCCGGCACGGACCCGTCCGAGGCCTTG 341

Qy      83 AGGTGCTGGAGGACCGCGC 101
           | | ||   | || ||||
Db     340 CGCTCCTCCCGCACGGCGC 322


RESULT 14
ABN17674
ID   ABN17674 standard; cDNA; 465 BP.
XX
AC   ABN17674;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:3825.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
```

```
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP01922.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 3825; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 465 BP; 85 A; 165 C; 139 G; 75 T; 1 other;

  Query Match             30.5%;  Score 30.8;  DB 24;  Length 465;
  Best Local Similarity   61.0%;  Pred. No. 38;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       4 ACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCC 63
           ||| | ||| |   |||    ||   || |||||| ||||||  |||  ||||  |||
Db     303 ACCCACCGCAGCAAGACCTGGCAGAACCTCGGTCGCGTGCAGGTGACCAACGGTTCGCTG 362

Qy      64 GGCGCGTTGCCGCGCCTCGAGG 85
             | |  ||||| ||  ||| |
Db     363 TTCTCCCTGCCGGGCGGCGACG 384


RESULT 15
ABQ90027/c
ID   ABQ90027 standard; DNA; 1899 BP.
XX
AC   ABQ90027;
XX
DT   01-OCT-2002  (first entry)
XX
DE   M. capsulatus gene #12 for DNA array.
XX
KW   Micro array; gene; ds; differential expression; gene expression.
XX
OS   Methylococcus capsulatus.
XX
PN   WO200255655-A2.
XX
PD   18-JUL-2002.
XX
PF   14-JAN-2002; 2002WO-NO00019.
XX
PR   12-JAN-2001; 2001NO-0000235.
PR   12-JAN-2001; 2001NO-0000239.
XX
PA   (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
PA   (TIGR-) TIGR.
XX
PI   Birkeland NK,  Eidhammer I,  Jonassen I,  Jensen HB,  Lien T;
PI   Lillehaug JR,  Lossius I,  Eisen JA,  Fraser CM,  Durkin AS;
PI   Salzberg SL;
XX
DR   WPI; 2002-557818/59.
XX
PT   Novel DNA array useful for determining differential expression of
PT   Methylococcus capsulatus genes, comprises polynucleotides or
PT   oligonucleotides representative for a selective number of Methylococcus
PT   capsulatus genes -
XX
PS   Claim 19; Page 52; 678pp; English.
XX
CC   The invention relates to a novel DNA array giving a representation of a
CC   number of Methylococcus capsulatus genes. The method of the invention is
CC   useful for determination of the differential expression of the genes of
CC   M. capsulatus, and for studying gene expression on a genomic scale and in
CC   gene expression assays of M. capsulatus genes. The sequences shown in
CC   ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
CC   invention.
XX
SQ   Sequence 1899 BP; 323 A; 580 C; 655 G; 341 T; 0 other;

  Query Match             30.5%;  Score 30.8;  DB 24;  Length 1899;
  Best Local Similarity   58.9%;  Pred. No. 36;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       4 ACCGAACGCCGATGGACGACCCACGCCCGCGGTCGGGTGCAGCCGACAGACGGCGCGCCC 63
           ||||   |  |  |  ||  |||   |  ||  | |||  |||  |  ||||||| ||||
Db     294 ACCGCGGGTGGCGGCTCGGACCAGAACGCCGTCCTGGTAGAGCAAATTGACGGCGAGCCC 235

Qy      64 GGCGCGTTGCCGCGCCTCGAGGTGCTGGAG 93
           |  || |  |  | || || ||| ||  | |||
Db     234 GTCGAGCTTCGGCTCCGCGACGTATTCGAG 205


Search completed: November  5, 2002, 13:38:47
Job time : 177.215 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_24950_25050
Perfect score: 101
Sequence: 1 cgtcgctcgggtgctgggct.....gatggcggtggagatccgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	34.4	34.1	372	12	BE769467	BE769467 QV1-FT003
2	34.4	34.1	372	12	BE771151	BE771151 RC5-FT007
3	34.4	34.1	378	12	BG027149	BG027149 602295826
4	34.4	34.1	393	10	BE018358	BE018358 bb79c01.y
5	34.4	34.1	427	12	BE818132	BE818132 RC6-BN027
c 6	34.4	34.1	430	12	BF765829	BF765829 IL2-CS005
7	34.4	34.1	484	12	BG122810	BG122810 602351767
8	34.4	34.1	499	13	BI089028	BI089028 602853578
9	34.4	34.1	502	10	AW250725	AW250725 2822540.5
10	34.4	34.1	505	12	BG252188	BG252188 602365058
11	34.4	34.1	510	10	BE408801	BE408801 601303319
12	34.4	34.1	545	10	BE294484	BE294484 601173412
13	34.4	34.1	554	13	BI084344	BI084344 602869861
14	34.4	34.1	575	12	BG120801	BG120801 602348564
15	34.4	34.1	610	13	BI334016	BI334016 602997113
16	34.4	34.1	616	10	BE295708	BE295708 601175750
17	34.4	34.1	619	10	BE266534	BE266534 601193521
18	34.4	34.1	629	10	BE278889	BE278889 601159188
19	34.4	34.1	638	10	BE278513	BE278513 601156011
20	34.4	34.1	646	12	BG575818	BG575818 602598623
21	34.4	34.1	661	10	BE407693	BE407693 601299712
22	34.4	34.1	663	12	BF795737	BF795737 602260334
23	34.4	34.1	687	10	AW411246	AW411246 fh11b12.x
24	34.4	34.1	692	12	BG767554	BG767554 602741706
25	34.4	34.1	709	10	BE264961	BE264961 601194065
26	34.4	34.1	723	13	BI251998	BI251998 602952639
27	34.4	34.1	724	12	BE732060	BE732060 601566626
28	34.4	34.1	758	12	BG697851	BG697851 602661063
29	34.4	34.1	768	12	BE797189	BE797189 601587818
30	34.4	34.1	769	12	BG764032	BG764032 602737036
31	34.4	34.1	771	12	BE744734	BE744734 601573128
32	34.4	34.1	784	12	BG697069	BG697069 602660278
33	34.4	34.1	801	12	BG697233	BG697233 602660470
34	34.4	34.1	814	12	BF308207	BF308207 601887455
35	34.4	34.1	814	14	BQ949497	BQ949497 AGENCOURT
36	34.4	34.1	822	12	BE873497	BE873497 601451412
37	34.4	34.1	828	12	BF315842	BF315842 601900959
38	34.4	34.1	836	12	BG281010	BG281010 602401462
39	34.4	34.1	840	12	BG740759	BG740759 602632731
40	34.4	34.1	846	12	BE799129	BE799129 601590094
41	34.4	34.1	847	12	BG675303	BG675303 602621606
42	34.4	34.1	856	12	BE795437	BE795437 601592995
43	34.4	34.1	873	12	BG697301	BG697301 602660553
44	34.4	34.1	877	12	BG281685	BG281685 602402375
45	34.4	34.1	878	14	BQ691636	BQ691636 AGENCOURT

ALIGNMENTS

RESULT 1
BE769467
LOCUS BE769467 372 bp mRNA linear EST 20-SEP-2000
DEFINITION QV1-FT0037-130600-241-c06 FT0037 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE769467
VERSION BE769467.1 GI:10223125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

```
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=QV1-FT0037-130
            600-241-c06&t3=2000-06-13&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 23
            High quality sequence stop: 372.
FEATURES              Location/Qualifiers
     source           1. .372
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="FT0037"
                      /dev_stage="Adult"
                      /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       66 a    110 c    130 g     66 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 372;
  Best Local Similarity   59.0%;  Pred. No. 9.6;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
             ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db     133 CGTGGCACACATTCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 192

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   | |||| | |||
Db     193 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 232


RESULT 2
BE771151
LOCUS       BE771151                373 bp    mRNA    linear   EST 20-SEP-2000
DEFINITION  RC5-FT0079-140600-031-H08 FT0079 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE771151
VERSION     BE771151.1  GI:10224809
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 372)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC5-FT0079-140
            600-031-H08&t3=2000-06-14&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 361.
FEATURES              Location/Qualifiers
     source           1. .372
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone_lib="FT0079"
                      /dev_stage="Adult"
                      /note="Organ: prostate_tumor; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
BASE COUNT       66 a    115 c    127 g     64 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 372;
  Best Local Similarity   59.0%;  Pred. No. 9.6;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
             ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db     185 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 244

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     245 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 284


RESULT 3
BG027149
LOCUS       BG027149                378 bp    mRNA    linear   EST 24-JAN-2001
DEFINITION  602295826F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4390230 5',
            mRNA sequence.
ACCESSION   BG027149
VERSION     BG027149.1  GI:12415488
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 378)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10079  row: d  column: 07
            High quality sequence stop: 378.
FEATURES              Location/Qualifiers
     source           1. .378
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4390230"
                      /clone_lib="NIH_MGC_86"
                      /tissue_type="osteosarcoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: bone; Vector: pCMV-SPORT6; Site_1: NotI;
                      Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                      Average insert size 1.533 kb. Library enriched for
                      full-length clones and constructed by Life Technologies.
                      Note: this is a NIH_MGC Library."
BASE COUNT       68 a    120 c    131 g     59 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 378;
```

```
  Best Local Similarity   59.0%;  Pred. No. 9.6;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db      48 CGTGGCACACATTCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 107

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     108 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 147


RESULT 4
BE018358
LOCUS       BE018358                393 bp    mRNA    linear   EST 06-JUN-2000
DEFINITION  bb79c01.y1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3048576 5'
            similar to gb:X13135 Murine mRNA fragment for fatty acid synthase
            (MOUSE);, mRNA sequence.
ACCESSION   BE018358
VERSION     BE018358.1  GI:8278381
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 393)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            image.llnl.gov/image/html/iresources.shtml
            Seq primer: -40RP from Gibco
            High quality sequence stop: 383.
FEATURES             Location/Qualifiers
     source          1. .393
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3048576"
                     /clone_lib="NIH_MGC_10"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
BASE COUNT       71 a    119 c    137 g     65 t      1 others
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 393;
  Best Local Similarity   59.0%;  Pred. No. 9.7;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db      67 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 126

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     127 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 166


RESULT 5
BE818132
LOCUS       BE818132                427 bp    mRNA    linear   EST 21-SEP-2000
DEFINITION  RC6-BN0276-160600-011-A12 BN0276 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE818132
VERSION     BE818132.1  GI:10250366
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=RC6-BN0276-160
            600-011-A12&t3=2000-06-16&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .427
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="BN0276"
                     /dev_stage="Adult"
                     /note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       83 a    129 c    144 g     71 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 427;
  Best Local Similarity   59.0%;  Pred. No. 9.9;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db     197 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 256

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     257 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 296


RESULT 6
BF765829/c
LOCUS       BF765829                430 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  IL2-CS0051-091100-223-D03 CS0051 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF765829
VERSION     BF765829.1  GI:12113729
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 430)
```

```
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-CS0051-
            091100-223-D03&t3=2000-11-09&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 368.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="CS0051"
                     /dev_stage="Adult"
                     /note="Organ: colon_est; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       72 a    135 c    147 g     76 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 430;
  Best Local Similarity   59.0%;  Pred. No. 9.9;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| |||| |   ||  || | | |  |  |      |
Db     417 CGTGGCACACATTCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 358

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     357 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 318


RESULT 7
BG122810
LOCUS       BG122810                484 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602351767F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4449808 5',
            mRNA sequence.
ACCESSION   BG122810
VERSION     BG122810.1  GI:12616319
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10234  row: f  column: 17
            High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4449808"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT       82 a    155 c    164 g     83 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 12;  Length 484;
  Best Local Similarity   59.0%;  Pred. No. 10;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| |||| |   ||  || | | |  |  |      |
Db     146 CGTGGCACACATTCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 205

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     206 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 245


RESULT 8
BI089028
LOCUS       BI089028                499 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602853578F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4995210 5',
            mRNA sequence.
ACCESSION   BI089028
VERSION     BI089028.1  GI:14507358
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11018  row: k  column: 19
            High quality sequence stop: 499.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4995210"
                     /clone_lib="NIH_MGC_10"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
```

Average insert size 1.5 kb. Library prepared by Life Technologies."
BASE COUNT 90 a 162 c 166 g 81 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 499;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
||| || | | |||||| ||| |||| | || || | | | | | |
Db 169 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 228

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
|| | |||||| |||||| | ||| |||||| | |||
Db 229 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 268

RESULT 9
AW250725
LOCUS AW250725 502 bp mRNA linear EST 07-JAN-2000
DEFINITION 2822540.5prime NIH_MGC_7 Homo sapiens cDNA clone IMAGE:2822540 5', mRNA sequence.
ACCESSION AW250725
VERSION AW250725.1 GI:6593718
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 502)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: 2822540.3prime
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Berkeley MGC sequencing project Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Base Calling ,' Quality Scores: PHRED from University of Washingtion Genome Center. Vector Trimming: cross_match from University of Washingtion Genome Center PHRAP suite. Poly-T Identification: patMatch.pl from Berkeley Drosophila Genome Project. University of Washingtion Genome Center: http://www.genome.washington.edu
Plate: LLCM9 row: K column: 21
High quality sequence stop: 379.
FEATURES Location/Qualifiers
source 1. .502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2822540"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 88 a 174 c 161 g 79 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 10; Length 502;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
||| || | | |||||| ||| |||| | || || | | | | | |
Db 38 CGTGGCACACATTCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 97

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
|| | |||||| |||||| | ||| |||||| | |||
Db 98 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 137

RESULT 10
BG252188
LOCUS BG252188 505 bp mRNA linear EST 13-FEB-2001
DEFINITION 602365058F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4473183 5', mRNA sequence.
ACCESSION BG252188
VERSION BG252188.1 GI:12762004
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10295 row: d column: 16
High quality sequence stop: 505.
FEATURES Location/Qualifiers
source 1. .505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4473183"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 92 a 162 c 169 g 82 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 12; Length 505;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
||| || | | |||||| ||| |||| | || || | | | | | |
Db 175 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 234

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
|| | |||||| |||||| | ||| |||||| | |||
Db 235 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 274

RESULT 11
BE408801
LOCUS BE408801 510 bp mRNA linear EST 21-JUL-2000
DEFINITION 601303319F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3637636 5', mRNA sequence.
ACCESSION BE408801
VERSION BE408801.1 GI:9345251

```
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM337  row: f  column: 05
            High quality sequence stop: 510.
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3637636"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT       91 a    157 c    175 g     87 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 510;
  Best Local Similarity   59.0%;  Pred. No. 10;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |       |
Db     286 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 345

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     346 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 385


RESULT 12
BE294484
LOCUS       BE294484                545 bp    mRNA    linear   EST 20-JUL-2000
DEFINITION  601173412F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:3528986 5',
            mRNA sequence.
ACCESSION   BE294484
VERSION     BE294484.1  GI:9178021
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM199  row: g  column: 03
            High quality sequence stop: 544.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3528986"
                     /clone_lib="NIH_MGC_17"
                     /tissue_type="rhabdomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      100 a    166 c    188 g     91 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 545;
  Best Local Similarity   59.0%;  Pred. No. 10;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |       |
Db     291 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 350

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db     351 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 390


RESULT 13
BI084344
LOCUS       BI084344                554 bp    mRNA    linear   EST 20-JUN-2001
DEFINITION  602869861F1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5014382 5',
            mRNA sequence.
ACCESSION   BI084344
VERSION     BI084344.1  GI:14502674
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1821  row: j  column: 15
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5014382"
                     /clone_lib="NIH_MGC_102"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
```

Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 101 a 172 c 194 g 87 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 554;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
Db 256 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 315

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
Db 316 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 355

RESULT 14
BG120801
LOCUS BG120801 575 bp mRNA linear EST 30-JAN-2001
DEFINITION 602348564F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4443518 5', mRNA sequence.
ACCESSION BG120801
VERSION BG120801.1 GI:12614310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10217 row: p column: 15
High quality sequence start: 2
High quality sequence stop: 575.
FEATURES Location/Qualifiers
source 1. .575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4443518"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 102 a 182 c 198 g 93 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 12; Length 575;
Best Local Similarity 59.0%; Pred. No. 10;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
Db 245 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 304

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
Db 305 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 344

RESULT 15
BI334016
LOCUS BI334016 610 bp mRNA linear EST 30-JUL-2001
DEFINITION 602997113F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5139530 5', mRNA sequence.
ACCESSION BI334016
VERSION BI334016.1 GI:15018660
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 610)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11343 row: i column: 03
High quality sequence start: 29
High quality sequence stop: 610.
FEATURES Location/Qualifiers
source 1. .610
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5139530"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.4 kb. Library prepared by Life Technologies."
BASE COUNT 112 a 190 c 207 g 101 t
ORIGIN

Query Match 34.1%; Score 34.4; DB 13; Length 610;
Best Local Similarity 59.0%; Pred. No. 11;
Matches 59; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
Db 280 CGTGGCACACATCCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 339

Qy 61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
Db 340 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 379

Search completed: November 5, 2002, 16:18:05
Job time : 1115.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_25850_25950
Perfect score: 101
Sequence: 1 gcgcgcgctttcgcccgacg.....ggtggtgctcaagcggctgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	68750	1	AF210843	AF210843 Sorangium
3		101	100.0	68750	6	AR193029	AR193029 Sequence
4		101	100.0	68750	6	AR199551	AR199551 Sequence
5		101	100.0	68750	6	AR199559	AR199559 Sequence
6		101	100.0	68750	6	AR199567	AR199567 Sequence
7		101	100.0	68750	6	AR201097	AR201097 Sequence
8		101	100.0	68750	6	AR208671	AR208671 Sequence
9		101	100.0	71989	6	AR172664	AR172664 Sequence
10		80.2	79.4	10910	1	AX024383	AX024383 Sequence
11		80.2	79.4	10910	6	AX024276	AX024276 Sequence
12		65.4	64.8	33529	6	AR166425	AR166425 Sequence
13		65.4	64.8	42603	1	AF188287	AF188287 Stigmatel
c 14		65.4	64.8	82746	1	AF453501	AF453501 Actinosyn
15		64.4	63.8	66808	1	SAU421825	AJ421825 Stigmatel
c 16		63.8	63.2	30000	6	AX250263	AX250263 Sequence
17		63.4	62.8	90348	1	AF497482	AF497482 Micromono
c 18		62.8	62.2	49736	1	AF319998	AF319998 Stigmatel
c 19		62.8	62.2	189050	1	AL646066	AL646066 Ralstonia
20		62.6	62.0	764	6	AX153716	AX153716 Sequence
c 21		62.2	61.6	27522	1	AB070942	AB070942 Streptomy
22		62	61.4	6564	6	AX069073	AX069073 Sequence
23		62	61.4	20679	1	MTCY19H9	Z83857 Mycobacteri
24		62	61.4	32870	1	AE007122	AE007122 Mycobacte
25		62	61.4	213050	1	AL646079	AL646079 Ralstonia
c 26		61.4	60.8	14634	1	AE007184	AE007184 Mycobacte
c 27		61.4	60.8	23740	1	MTV026	AL022076 Mycobacte
28		61.2	60.6	758	6	AX153712	AX153712 Sequence
29		61.2	60.6	5088	6	AX153795	AX153795 Sequence
30		61.2	60.6	34071	6	AX153790	AX153790 Sequence
c 31		61.2	60.6	42717	6	AX153789	AX153789 Sequence
c 32		60.8	60.2	333	6	AR128456	AR128456 Sequence
33		60.8	60.2	5481	6	AX069075	AX069075 Sequence
34		60.6	60.0	3573	1	AB016763	AB016763 Streptomy
35		60.2	59.6	3792	6	AX195972	AX195972 Sequence
36		60.2	59.6	109519	6	AX195929	AX195929 Sequence
37		60	59.4	11219	1	SERERYAA	M63676 Saccharopol
38		60	59.4	11219	6	AR049367	AR049367 Sequence
39		60	59.4	11219	6	AR095528	AR095528 Sequence
40		59.6	59.0	671	6	AX153719	AX153719 Sequence
41		59.6	59.0	49377	6	I88042	I88042 Sequence 1
42		59.6	59.0	67523	1	SCU24241	U24241 Sorangium c
c 43		59.4	58.8	19601	1	AB070944	AB070944 Streptomy
44		58.4	57.8	3185	1	AF411573S4	AF411576 Actinomad
45		58.4	57.8	30000	6	AX250263	AX250263 Sequence

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"

CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"

CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLRPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"

gene 1998. .6263
/gene="epoA"

CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYERPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"

gene 6260. .10492
/gene="epoB"

CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"

gene 10639. .16137
/gene="epoC"

CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRRLEDDWFLELEWEP
AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"

gene 16134. .37907
/gene="epoD"

CDS 16134. .37907
/gene="epoD"
/note="type I polyketide synthase; epothilone PKS modules 3, 4, 5, and 6"
/codon_start=1
/transl_table=11
/product="epoD"
/protein_id="AAF62883.1"
/db_xref="GI:7453560"
/translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

Query Match 100.0%; Score 101; DB 1; Length 58733;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25850 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 25909

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25910 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 25950
```

RESULT 2
AF210843
LOCUS AF210843 68750 bp DNA linear BCT 21-JAN-2000
DEFINITION Sorangium cellulosum strain So ce90 epothilone biosynthesis gene cluster, complete sequence.
ACCESSION AF210843
VERSION AF210843.1 GI:6724237
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 68750)
AUTHORS Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M., Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D., Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
TITLE The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90
JOURNAL Chem. Biol. 7 (2), 97-109 (2000)
MEDLINE 20130945
PUBMED 10662695
REFERENCE 2 (bases 1 to 68750)
AUTHORS Molnar,I.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1999) Natural Product Genetics, Novartis Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box 12257, Research Triangle Park, NC 27709, USA
FEATURES Location/Qualifiers
source 1. .68750
/organism="Polyangium cellulosum"
/strain="So ce90"
/db_xref="taxon:56"

CDS <1. .1826
/note="ORF1"
/codon_start=3
/transl_table=11
/product="unknown"
/protein_id="AAF26904.1"
/db_xref="GI:6724238"
/translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"

CDS complement(1900. .3171)
/note="ORF2; similar to hypothetical proteins from Mycobacterium tuberculosis and Streptomyces coelicolor and to DD-peptidases"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26905.1"
/db_xref="GI:6724239"
/translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
IGIVMTQSAGFLFSGALERFWRSVYVATESA"

CDS 3415. .5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
YLGSTVESVVVRSRVPVALLVAHGGTREQVR"

CDS complement(5612. .5992)
/note="ORF4"
/codon_start=1
/transl_table=11

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MEHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MALRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRFVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPFESGRRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251. .21749
                     /gene="epoB"
     CDS             16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match           100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
```

```
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 3.1e-11;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  25850 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 25909

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25910 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 25950


RESULT 10
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match            79.4%;  Score 80.2;  DB 1;  Length 10910;
  Best Local Similarity  87.1%;  Pred. No. 5.8e-07;
  Matches   88;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          |||   |||||||| || ||||||||||||||||||||||||  || |||||||  ||
Db   2897 GCGTTTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGGCTTTGC 2956

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          ||||| |||||| |||||||||||||||||||||||||| |
Db   2957 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGCTCC 2997


RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match            79.4%;  Score 80.2;  DB 6;  Length 10910;
  Best Local Similarity  87.1%;  Pred. No. 5.8e-07;
  Matches   88;  Conservative    0;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          |||   |||||||| || ||||||||||||||||||||||||  || |||||||  ||
Db   2897 GCGTTTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGGCTTTGC 2956

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          ||||| |||||| |||||||||||||||||||||||||| |
Db   2957 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGCTCC 2997


RESULT 12
AR166425
LOCUS       AR166425               33529 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 3 from patent US 6280999.
ACCESSION   AR166425
VERSION     AR166425.1  GI:16241741
```

```
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33529)
  AUTHORS   Gustafsson,C., Betlach,M.C., Ashley,G., Julien,B. and Ziermann,R.
  TITLE     Sorangium polyketide synthases and encoding DNA therefor
  JOURNAL   Patent: US 6280999-A 3 28-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .33529
                     /organism="unknown"
BASE COUNT     4489 a   9518 c  14470 g   5046 t      6 others
ORIGIN

  Query Match            64.8%;  Score 65.4;  DB 6;  Length 33529;
  Best Local Similarity  78.8%;  Pred. No. 0.00037;
  Matches   78;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||| ||| | |||||||||||| |||||||||||||||||||    |  || ||||||||||
Db   10351 GCGTGCGGTCTCGCCCGACGGACGGTGCAAGACGTTCTCGCAGGCGGCCGACGGCTACGG 10410

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             || ||||||||   ||  |||||||||| | ||| ||
Db   10411 CCGCGGCGAGGGCGTCGGGGTGGTGGTGCTGATGCGTCT 10449


RESULT 13
AF188287
LOCUS       AF188287               42603 bp    DNA     linear   BCT 12-JAN-2000
DEFINITION  Stigmatella aurantiaca myxothiazol gene cluster, complete sequence.
ACCESSION   AF188287
VERSION     AF188287.1  GI:6635393
KEYWORDS    .
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 42603)
  AUTHORS   Silakowski,B., Schairer,H.U., Ehret,H., Kunze,B., Weinig,S.,
            Nordsiek,G., Brandt,P., Blocker,H., Hofle,G., Beyer,S. and
            Muller,R.
  TITLE     New lessons for combinatorial biosynthesis from myxobacteria. The
            myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca
            DW4/3-1
  JOURNAL   J. Biol. Chem. 274 (52), 37391-37399 (1999)
  MEDLINE   20069734
   PUBMED   10601310
REFERENCE   2  (bases 1 to 42603)
  AUTHORS   Silakowski,B., Schairer,H.U., Ehret,H., Kunze,B., Weinig,S.,
            Nordsiek,G., Brandt,P., Bloecker,H., Hoefle,G., Beyer,S. and
            Mueller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1999) NBI/MX, GBF, Mascheroder Weg 1,
            Braunschweig 38124, Germany
FEATURES             Location/Qualifiers
     source          1. .42603
                     /organism="Stigmatella aurantiaca"
                     /strain="DW4/3-1"
                     /db_xref="taxon:41"
     gene            1. .834
                     /gene="mtaA"
     CDS             1. .834
                     /gene="mtaA"
                     /note="Ppant transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaA"
                     /protein_id="AAF19809.1"
                     /db_xref="GI:6635394"
                     /translation="MPTSSPALPLLKLPPDEVHVWIVEPERITEPGLLESYRALLDPG
                     ERDKQQRFYFERHRLQYLVSHALVRLTLSRYAPVAPEAWSFSANQYGRPEIRGEEKPW
                     LRFNLSHTDGMALCAVARDVDVGADVEDTERRGETVEIADSFFAPAEVASLRALPVSG
                     QRERFFDYWTLKEAYIKARGMGLSLPLDQFAFEVSQGLSTRISFDPRLVDEPSQWQFV
                     RFRPSQRHAAALAVRRPSEAPLTVRFQRTVPLQDDAPAEYLSRERIQPLRLRMPGVGG
                     G"
     gene            1388. .13399
                     /gene="mtaB"
                     /note="PKS"
     CDS             1388. .13399
                     /gene="mtaB"
                     /note="PKS; MtaB loading domain and module 1: polyketide
                     synthase domains: ACP, KS, AT, AT, DH, ER, KR, ACP; MtaB
                     module 2: polyketide synthase domains: KS, AT, DH, KR,
                     ACP"
                     /codon_start=1
                     /transl_table=11
                     /product="MtaB"
                     /protein_id="AAF19810.1"
                     /db_xref="GI:6635395"
                     /translation="MSTEHGVGGVGRSTGQLRDWLVQQVASQVGLEAHTVRITEPFFS
                     LGVGSVFATRLIAELARWLERPLSPTLVWEHPTIDKFARHLSGAEPPRAEMHAPGPEA
                     AVDGAEPIAIVGLACRFPKAPDAKAFWRVLQEGRDCVTEVPSNRWDASLYYDPAPGTP
                     GKIVTRYGAFLDSVDRFDPVFFGISPREAGQLDPQQRLMLELSWEALEDAGIAPRVLR
                     GSRTGVFTGMVWRDYADLQTIHNAAPTLHTCTGYAGGIIANRISYALGLEGPSLAIDT
                     ACSSSLVAIHEACRSLRSGESSLALAGGVQLNLSPRTFMAVSAFGALAPDGRCRTFDA
                     RASGYVRGEGGGVLVLKPLSRALADGDAIYCLIRGSAVNNDGASNGLTAPNPKAQERL
                     LRDAYAVAGMDPAQVHYVELHGTGTPLGDPIEAGALGAVLGAARE AERPLHVGSVKTN
                     IGHLEAAAGIASVIKVALSIRHQQLPASLHFEQPNPRIPFDDLRLRVQTQLAPWPAEK
                     DPAVAGISAFGFGGTNCHVVVEEHRGGSRTAAVPVSTPHAQSPRHPKTAFLFAGAGSA
                     WWGMGRELLGAEPVFRARFEECDAAFRPLGGWSLVEALLAPRVRSRFEQGERQLPLVL
                     AFQISLAALWRSWGIEPSATVGYSVGEWAAAQVAGILTIEEVFALAYHYIQVQKGVAG
                     GVGMAVVGLSGAEVAERIAPYGGRISIASENGPRSTGLGGEAAALQDFTASMTAAGIF
                     ARLVEIDFYAHIPQVEALKPLFLQSCPVLRPSPGRVPMVSTVSGAFVDGGTLGAEYWA
                     SNFREPVRFATAMRTLLEGGYDVVLDVNPHPIHEHSVLENVRAGTASAEFIASMRRGE
                     PSVEGLHRALTQLSQAGGPVRWERVVPGTGAGSSAPVRTELFVASARTEQALREQADR
                     IAEQLRRDAASSLTDLCYTAGARRSHHEHRLALVAETRAALEEGLAAFARGEIRPNPS
                     QGMVRPGEAPRLGFVFSGQGSFPLGSSVQLFEEEPSFRKAFERSERLFRAQTGASLIE
                     ALHAERESQHPPRTEIEQPLLFAFQVALAALWESWGIRPDAVVGHSVGEISAAHVAGV
                     LSLEEALSLVVSRSRLMGRAAGQGGMATVEASPARIEPLLGKCGTDLAIAAVNGPENV
                     VVSGALAALASLRDELTGQGLRVRVLNVNYAFHSAQMAPYLGDLDPVLRELKPKPPRL
                     PIYSTVHGRLASASDFGTDHWRANVRERVLFAPAVEAMRADGHGAFVEISPHPVLLQP
                     LSRILDASPASGASALFSTRRDRAPRAALLGNLGTLYTLGHEPSWKGVFPDLGRTVEL
                     PSYPWQRQRCWFDAAPQPLAEPVHPLLGRRVDAPPGQTVFEATFRSDRLQLLSDHRLY
                     GTVVVAGASMLGMMLVAARQLWGEAPCALQDVLFLQPLVLGDTEDRLVQVSIATDGAF
                     RILSWAGQGGEGTSSHWIEHATGRLERGGAEETEAGREVLSEVEARCRPLASAESFYE
                     GIARQGLGLGRQFRWLEALWQGDREAFGRLRVPETDDGAEDYRFHPGLIDSCFQVLGA
                     VAGFASIEGLAFVPIGVGRFALRGEARPPRFCHLVLRRGSGKANEASSVDLRLLDESG
                     LVVAEVSGLVIRRAPREVMLRRDVTAPPLVLRTEWQPQMLSGASQAGPEAWLLLADRG
                     GLAARLAEQLSSRGSRVEVLAAEENAQSMLGRIEAIDASGLACAGVVDLRGLDEGSSS
                     DAVQGAVKRVQQLVTRRAAEALRLVLVTRGTQVTGSQDGSIAFEQAPLWGLGNVIALE
                     HPELRCQRIDLDPSPGAPDAAQLAAELAAPGREDRVAFREGRRLVARLVRTKLPPAVR
                     ALPLEGDATYLIVGGLGGIGLELARWMVGRGARHLVLMGRSAPSAHAQETLAALRDAG
                     ATISVALGNVAEREALARVLGDIRAAMPPLRGVMHAAGVLDDGLLVRQSEERLARVLA
                     PKIQGAMNLHRETVGDRLDFFVLFSSVSSVFGTSGQAGYASANAFLDALALHRRAQGL
                     TALSINWGPWADVGMAARLPPEAGKSWSSFGLGLLAPSAGLAALERLLSSEAAQVVVA
                     GASEAEAKGDFAPFFDALSSRKPEASRPASVGGALVGASPRELRARLLSLLQVQAATV
                     LGLPPSAPIDPRRPLQELGLDSLMALQFRNSLAASLGRSLPATLLFDHPTLEKLFEFL
                     ARDLTPASEASPAKSRRPVKLPARDAEGAEPIAVVGMACRLPGSVRSPEEFWRLLANG
                     VNAVREVPKDRWSLDTWYDPDPETPGKMYSRHGGFLDEVDPFDASFFGISAPEARSMD
                     PQQRILLELGWEALESAGIAVDQLKGSPTGVFLGICLSDYALLELNSPDPTGINAYSG
                     SGGVLSVAAGRIAYALGLEGPTLAVDTACSSSLVAAHLACQSLRNGECDLALVGGSNL
                     LLSPRMSVYFSKLKALSPDGACKAFDSSANGYVRGEGAGVLVLKRLSEALADGDPILS
                     VMRGTAVNQDGRSNGLTAPSRAAQERVIESALRQGGVSPLDVSYVEAHGTGTSLGDPI
                     EVQALAAVLGKDRPRSAPLGIGSVKTNVGHLEGAAGVTSLIKAILALRHRELPRSLHF
                     KTPSPHIPWDELPVEVISEHRPWRVPEGKRRILGVSSFGFSGTNAHVILEEAPEAGPR
                     SKAPSIPERPELLVLSARAPQALQEAAVRFSAQLSGRAGGDAAMLHDVCYSANCRRSH
                     HEHRLAVVARSRTEAAEALEAFLRGLAHPDVQAGYALPGAVPRVVFLFTGQGGQWPGT
                     GRALLEEEPVFRSALQAADEVLKGLAGWSIIEELQRETPRFDQAAISQPAHFALQSAL
                     LALLRSWGIEPDAVAGSSIGEVAAAYAAGVLTLEEAMRITVARSRILHAVAGKGAMAL
                     VELSEEEATTEVAAYAQRLSVAAINSQNSTLLSGEAGAMDELLGALTARGVSCRRIGL
                     DYASHSPAVEPYIDELKQALQGLRPQPAAIPIFSTVTGRLAEAGDFNSDYWARNLRQP
                     VRLAPVLTTLAREDYTLFVELGPNAVMARPVQQCFQAVGKSATLLSTLRRGESERHTL
                     RSLVGGLFIAGVTPAWNRLFPEGGSLVSLPSYPWQKERYWIASPTSARLQLPAPSQGV
                     LHGRRVRSPLAESLFETDLSVEKMPFLREHLLHGALVVAGAAHVALALSALRELRGTA
                     PVRLADIVFPRPLVLKEGVENPVSVVVGREQGGRSALRIQSPQGSEPEWVTHAEGWVL
```

PASSQVEGPAQVALEELRARCPNGLEVSPVYKAMDAEGIGLGRRFQWVEQLWQGDGEA
LGRMRAAESADQHPSAPLHPGLIDSIFHVLRAVPGIGRGDGVPYIPIGIEGLCFLAPP
RGVLWAHVKLRPEGSFDGETLSADFRLFEESGKTVAWIDGLIVKRAPREAFQRHEAQD
LSEKLFQVRWQSMPLAPAVPERGPWLLVADRGPAGEELAARLKREGQVCHVLAAPESA
VTGEALLQRVEEALALEAYRGVVYLSALEAEQESPQGKSQAALEGVSGAVQVLQAIAK
LERGTPPRLWLVTRGAQALEESPRPLALAQAPLWGLGLAVRDEHPELRCTLIDLDPDP
AADGVGCLFQELLTPDREDRIAWRGGARFAQRLSFLEAVPEPGTAPISPDGTYLITGG
LGGIGLRCARWLVDAGARHLALLGRSGPSEAARSVLGALEAEGVTVRVLQADVAVRED
VSKVLSALRVDMPPLRGVLHAAGLLDDALLVRQDRDRIQRVLQPKVAGALNLDLETRG
DPLDFFVLFSSIAAMLGGEGQGNYSAANAFLDALAFARRAQGLPAHSLAWGPWAEVGM
VAARSERSAHLAALGIQSIPPDEGVAILGRLLRAASAHVGIFQIDWTRWQQRHAEGHA
PLLAGILKQQAAAPSQDSSEEAPDIQELIRATPEERQKRAEAFLRREVTRILDLPRPL
ADGRLPLVRFGFDSLMSLKLKLKARLESTLGLSISAAKLLSGLSLDKLVALVLERIETLQ
PTPVSTDDQMEDFRF"
gene 13537. .17409
/gene="mtaC"
CDS 13537. .17409
/gene="mtaC"
/note="peptide synthetase; peptide synthetase domains: Heterocyclisation, adenylation, PCP note: motif with weak homology to putative NAD binding proteins from Pyrococcus species located behind PCP; this motif may be involved in thiazole formation from the proposed thiazoline intermediate of the biosynthesis of myxothiazol"
/codon_start=1
/transl_table=11
/product="MtaC"
/protein_id="AAF19811.1"
/db_xref="GI:6635396"
/translation="MIELLRPKEAVPDAARQFEPFPLTDIQETYFVGRGTDFGISGMS
CHLYHELDTRGMDVARLSRAWQRLIEQHVMLRSVVLPTGEQRVLESVPPFSIPVLDLR
GQAPALVETKLAEIRQQLSTRSTPPGQWPTFELRATLLDEGRTRIHFDFDAITADASA
VLTLLEEWREFYLSPDRFLKQVPVSFRDHVLAEVAARDTDSFRQAEAYWNARLNDMPD
APALPVATEPERLMRPSFRRLSGRLDAARWARLKERAANAGLTASGVVCAAFAEVLGA
WSESPRFTLNLTLFRRLPIHPEVDRLLGDFTTNVLLEVDRSGESFTARAVRLRDQLAE
DLEHTHFSGVRVMRQRAQMKKGNGAIMPVVFTSLLGHRSRVPGGGSPFSWLGEVVYAI
TQTPQMWLDHQVREEAGALYCHWDSPEGLFPEGLLEGAFEAYTDLLSRLADDNALWEE
GTVVRLPEAQRARRETFNATTASIPEARLDGLFLAQAEQGPQRMAILEPALRMSYGEL
LGQAGSLAAHLAELGARPNELTAVVLEKGWRQIVAVTAIHLAGSAYLPIDPALPEERR
RLLLEEGQVKVVLTESHLAERLSWPASVRVVAIADTARGAPPRLPARAASDMAYCIYT
SGSTGRPKGVKIEHRAAANTILDINERFGVGPADRVFALSSLGFDLSVYDIFGSLAAG
AAIVLPQPEVTRDPAHWLQVLRQQGVTIWNSVPTLMEMLVDQVEARGELLPETLRLVM
LSGDWIPVTLPDRIRRLSRDIRVISLGGATEASIWSIIHPIGRVDPAARSIPYGRPML
NQRFYVLDEKLAPRPDHVPGDLYIGGVGLAQGYFRDEETTRARFFNHPRTGERLYRTG
DLGRFLPEGEIEFLGRKDTQVKIAGHRIELGEIEAVLQRHPSIREAVVAAPGERTSRR
LVAYLVPVEGQEAPSDEVLRSFLAAALPQYMVPAVFVRLEHLPLSSNGKVDRKALPEP
VLQPSRASGLTDARGGKILSAIMRLVSEVLKRPSIEPEAGLLQLGATSVELIKLSMML
EQEFGSRPTITELLSLRNLSEIAAYYVERQPAEEARKSDEAELVPAGIRRFDAGAVTA
QLEAPPEDEARRQCYFKRVSHRTFSAEPVEARVLARLLSCLSPLKVEGQLKYQYGSAG
GIYGVQTYLSIATGRVRGLDAGAYYYDPLRHTLVQLGGASALPAALDSSANAAMADSA
AFSLYLVGDRRVISQRYGERWRDLALIEAGLMAQLLETRAAESDLGLCQLEELRFEGL
SKALRLEEEHMYLHGLVGGGLAWEEGAL"
gene 17406. .27281
/gene="mtaD"
CDS 17406. .27281
/gene="mtaD"
/note="PKS/NRPS; MtaD module 1: peptide synthetase domains: heterocyclisation, adenylation, PCP note: motif with weak homology to putative NAD binding proteins from Pyrococcus species located between adenylation core regions A9 and A10; this motif may be involved in thiazole formation from the proposed thiazoline intermediate of the biosynthesis of myxothiazol; similar to the motif found in MtaC; MtaD module 2: polyketide synthase domains: KS, AT, DH, KR, ACP"
/codon_start=1
/transl_table=11
/product="MtaD"
/protein_id="AAF19812.1"
/db_xref="GI:6635397"
/translation="MSTSEFIVYLQQKGIELWIEGDTLRFRAPPGHLTPELRTALKDK
KEQLLEHLRSASRSANLQTVASPPDAASRFAPFPLTDIQNAYWVGRQGAFDAGGVAAH
SYLELAFDTLDLPRLESILQRLIEYHDMLRMVVLPTGEQVVLEKVPPFTLTAYDLAGA

PASQIEAHFEAIRGELSHQVLPTDRFPLFDVRASKLAGGRVHLHISFDLLTADAFSVQ
ILIEQCASLYLNPEAQLVPFKRTFRDYCVLAAQRADPSSRAYQNSLAYWRNRLEKLPN
APDLPLVNDTDVKGQHQFVRRKAGLDQASWRAFRNHAKAAGITASMALCAAYGETLRA
YSRSNRLTLNLTLFNRLPLIEDVERIVGDFTSGILLELDGSKPESFAARARRLQGQFI
NDLEHSAVSSVKVLREAIRLGRWDAGTAMPYVFTSLISETGRSLSVSEGVKIVDVISQ
TPQVWLDHQVFELDGGIYYSWDSVDALFPAGLIDAVFESYGRLLHRLSSDASAWDALS
RQPLLPAQLASRAEYNATERPVSQERIESLFLRQAERQPGAVAVIAGEQVLSYGKLER
RSARIAAWLIAQGVKPDQLVAIVAEKGVEQVVAALAILRAGAAYLPLSPSLPPERLHA
LLEEAQADVVLTQSALESSLRWPEGPRRLSVDRDEFLEAPKGALPPVQGNNLAYVIYT
SGSTGRPKGVMIDHRGAVNTLLDMNERFNVGPRDRVLALSSLSFDLSVYDLFGVLAAG

Query Match 64.8%; Score 65.4; DB 1; Length 42603;
Best Local Similarity 78.8%; Pred. No. 0.00036;
Matches 78; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
Db 28101 GCGCTCGCAGTCGCCTCAGGGGCGCTGCAAGACCTTCGACGCCTCCGCGGATGGCTACGT 28160

Qy 61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
Db 28161 CCGGGGAGAGGGCTGTGGCGTCGTGGTGCTCAAGCGGCT 28199

RESULT 14
AF453501/c
LOCUS AF453501 82746 bp DNA linear BCT 18-JUN-2002
DEFINITION Actinosynnema pretiosum subsp. auranticum maytansinoid antitumor agent ansamitocin biosynthetic gene cluster I, partial sequence.
ACCESSION AF453501
VERSION AF453501.1 GI:21449342
KEYWORDS .
SOURCE Actinosynnema pretiosum subsp. auranticum.
ORGANISM Actinosynnema pretiosum subsp. auranticum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Actinosynnemataceae; Actinosynnema.
REFERENCE 1 (bases 1 to 82746)
AUTHORS Yu,T.W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q., Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
TITLE The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from Actinosynnemapretiosum
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 7968-7973 (2002)
MEDLINE 22056096
PUBMED 12060743
REFERENCE 2 (bases 1 to 82746)
AUTHORS Yu,T.-W., Bai,L., Clade,D., Hoffmann,D., Toelzer,S., Trinh,K.Q., Xu,J., Moss,S.J., Leistner,E. and Floss,H.G.
TITLE Direct Submission
JOURNAL Submitted (28-NOV-2001) Department of Chemistry, University of Washington, Box 351700, Seattle, WA 98195-1700, USA
FEATURES Location/Qualifiers
source 1. .82746
/organism="Actinosynnema pretiosum subsp. auranticum"
/strain="ATCC 31565"
/sub_species="auranticum"
/db_xref="ATCC:31565"
/db_xref="taxon:42198"
misc_feature <1. .82653
/note="cluster I of the maytansinoid antitumor agent ansamitocin biosynthetic gene cluster"
gene <1. .1678
/gene="asm37"
CDS <1. .1678
/gene="asm37"
/codon_start=2
/transl_table=11
/product="unknown"
/protein_id="AAM54115.1"
/db_xref="GI:21449383"
/translation="IVPGLVVALAATGVLAAPASAAITSPTVTSTATTVTFEVGRTGT PAFSRVYVDTDRSAATGYAQGGIGADHLLENGSLYRHSGGGWAWTQVRTVPFTSASGV ARWTVERADLGEQASPNDVDVVFQVEAPVETSAKVTHTYTAGAGGSVAYQPSSANVPN

```
                PERGLYHHTGDCDKADLSADALRRYRTEQGITLAMCVFYLAEHRSSPIAQTALDRLQR
                QFDAARVAGVKLVLRFAYTASESGEDAPKERVLAHIDQLAPLLRANADVTAVVQAGFV
                GAWGEWYYTKNFGNAGVVSATDQANRKAVVDKLLTALPADRAVQLRTPKFKRALYGGG
                STRIGHHNDCFLASDTDFGTYEDPATEYPYLEAETRTLPMGGETCAVNPPRTDCPTAT
                AELSRFHWTYLNTDYQQDVLRAWSAQGCLAEVTAKLGYRFTLRDSVFPTTTTRGAALP
                VRVNLTNDGYAAAVNQRTVNLVLRNTTTGALTRLPLTADPRTWLPGARAIAQDAQIPP
                SLATGQYALLLELADPALPGRPEYSIQTANTGLWDARTGLNDLKQTITVS"
gene            1825. .2142
                /gene="asm36"
CDS             1825. .2142
                /gene="asm36"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAM54114.1"
                /db_xref="GI:21449382"
                /translation="MVTASAVLCARAGCGAPRPAGATARWRFCSTACKQAAYRLRRKA
                GSASHEHEDEQLREALLALVARARTLTEATPPWSAQATALVGLARRVVELAAARDAGV
                SGG"
gene            complement(2129. .3307)
                /gene="asm35"
CDS             complement(2129. .3307)
                /gene="asm35"
                /codon_start=1
                /transl_table=11
                /product="bicyclomycin resistant protein"
                /protein_id="AAM54113.1"
                /db_xref="GI:21449381"
                /translation="MAVGKADSAVGGWLPALLIVALTGTIIEIDMSVPSFPDIARELG
                VGGPAVQLTVTLNFLGYCLGALCHGPLSDRFGRRPVLLGGSAVMLVGALGCALAPTIG
                LLLTARFVQGIGASAAAVLVFVIIGDRYRGPAALRMYGLVNAAMSALMTCAPLLGGAI
                NRVAGWRGNYLAVCATTALSLLLVFLFLPETRTRARTAITARSVLADYRRLLGSRAYL
                AASLTPSLLFAAYLVFIAASPFLYTAAFGLGVTGFALHLLVVVASFAVPSAFATGIIP
                LLGGPAGAVRWGLGAVVVGVLGFLLLGDGPVAVTGSVALFCVGFAVCYPVVFDRSMSV
                FPELAGAASSATMGLRALLVTLLTALAGALSTGGPFAPAAVMAGAVAAVVLLSALSPR
                "
gene            complement(3450. .4064)
                /gene="asm34"
CDS             complement(3450. .4064)
                /gene="asm34"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator"
                /protein_id="AAM54112.1"
                /db_xref="GI:21449380"
                /translation="MARRNPERRAALLDAAVEVLAGEGARGLTFRAVDQRAGVPAGTA
                SNYFAGRAEILTGAGERVYQRLLDQAPLGETPDGPRDRSRVTELMHALVDRVSAFPTG
                FLALLELRLEAARNPELREVLTARVRADLRFNAEHHAASGLPGDGTTVALLWLALNWL
                ILERLTLPDLLTDHQRRDLVTALVDRLLAGQPEEAGSPFDESGK"
gene            4151. .4723
                /gene="asm33"
CDS             4151. .4723
                /gene="asm33"
                /codon_start=1
                /transl_table=11
                /product="dihydrofolate reductase"
                /protein_id="AAM54111.1"
                /db_xref="GI:21449379"
                /translation="MRKLVYFIAASIDGVIAGPEGEHDFYPTAEPMVEHLRAEYPETT
                PTHVRPLVGMPLDTPNRRYDTVLMGRGSYQPALDAGVGSPYAHLRQYVVSRSLPDADP
                AVALVRADPVGLVRELKREDGLDVWLCGGADLAGQLLPEIDELVVKTYPVVAGGGTPL
                FRAGFAPRAFRPVDVRAFDHGGVVTTYRPA"
gene            complement(4854. .5057)
                /gene="asm32"
CDS             complement(4854. .5057)
                /gene="asm32"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAM54110.1"
                /db_xref="GI:21449378"
                /translation="MWITEQNARIPARRSSGRSNARRCAVRWPRPQPRSATRAPGPRS
                SANADSRARSRGRDPRTSLNRSA"
gene            5415. .6461
                /gene="asm31"
CDS             5415. .6461
                /gene="asm31"
                /codon_start=1
                /transl_table=11
                /product="RNA polymerase sigma factor"
                /protein_id="AAM54109.1"
                /db_xref="GI:21449377"
                /translation="MWIGYSADGRRRRPVRWWGVTEAFTWETDEDIKLAATLRRDAEA
                ANAADSTRAYLKAIGRVRLLTADQEVELARRVEAGLYAEHLLDLPPAARGSRRPVELR
                RVVRDGQRAKDHLLEANLRLVVSIAKRYTGKGMALLDVIQEGNLGLVRAVEKFDYAKG
                FKFSTYATWWIRQAITRALADQSRTIRLPVHMVEVVTKLARIERDLLHRNGREATAEE
                VAVELDVTPEKVVELRQHARTPVSLDQTIGAEADTSIGDLIADENAVAPLEAVSSALL
                SADLRAVLETLTPREAGVVRMRYGFPDGVLHTLDEVGRVYGVTRERIRQIESRTMAKL
                QHPARSRGLEGYLD"
gene            complement(6501. .9518)
                /gene="asm30"
CDS             complement(6501. .9518)
                /gene="asm30"
                /codon_start=1
                /transl_table=11
                /product="cytochrome P450"
                /protein_id="AAM54108.1"
                /db_xref="GI:21449376"
                /translation="MVATGTRIPGPKPLPLVGNLLDVLTSDLDTDVDFLDRCHREHGG
                IVALTFAGQRQVFASSHELVARMCSDPSWGKAVHPALEQVRDFAGDGLFTARGDEPNW
                GKAHRLLMPAFGPTAMRDHFPAMLDIAEQMLVRWRRFGPDHRIDVADDMTRLTLDTIA
                LCAFGARFNSFYRDRAHPFVDAMVRSLVEAGERAERLPGVQPFLVGRNQRYRDDIATM
                NRIADGIVAARAALPAGERPDDLLERMLTCADPVTGERLSARNVRYQLATFLIAGHET
                TSGLLSFAVHRLLAHPEVLRKAKDAVDGVLGDRVPAFEDLARLDYLGQVLRETLRLHP
                TAPAFALAPDEPAELGGHAIGAGEPVLVMLPTLHRDPAVWRDPDVFDPERFAPERMDE
                IPACAWMPFGHGARACIGRPFALQEATLVLALVLQRFDLALADPDHRLTIKQTLTLKP
                DSLVVRARPRADRPGATATVETVVPHQVPATHRHGTPLHVFYGSNGGSGEGLARTIAG
                DGAARGWATSVAPLDDAVRALPASGPVVIVSSSYNGAPPDNAAHFVRWLTQDGPDLSG
                VDYLVLGCGNLDWSATYQRVPTLIDEAMAAAGARRLRERGATDARADFFGDWERWYEP
                LWPLLSAECGVEVGEIGPRFRVVESDAADGLGDLASAVVLENRELVRGPDAGSKRHLE
                LRLPDGTSYRTGDYLSVLPQNHPDLVRRAVARLGTRAERVVTVESSAPTGLVPVGRAL
                RVDELLTRCVDLSAPAGAGVVARLAERCPCPPERAELAATTGATLLELLERFPSCAVD
                LALALELLPAPRTRLYSISSAAEEQRAEVALTVSVTGVTSGYLSRVRPGDRVAVGIAS
                PPESFRPPADNTVPVVLIAAGTGIAPFRGFLRARAALGGEPGPALLLFGCRGPELDDL
                YAEEFAALGDWLEVDRAYSRHPDGEVRHVQHRLWQRRDRVRELVDAGARVYLCGDATR
                VGPAVEEVLGRIGPGAGWLDALRAGGRYATDVF"
gene            9664. .10245
                /gene="asm29"
CDS             9664. .10245
                /gene="asm29"
                /codon_start=1
                /transl_table=11
                /product="transcriptional regulator"
                /protein_id="AAM54107.1"
                /db_xref="GI:21449375"
                /translation="MSAAVRREQLVAAALRVMKRDGIAAATTRAICAEADMPHGAFHY
                CFRSKQELYTALLTTGLNTSLDAAWPAISPDLPAVDNVRTMLAAFWTAFQADPEAELV
                LGELVTLAQREPELRELPAWQHREYVAKAVAHLDRLAEEAGFDYSVDRALLAELVVSV
                LDGVASSWFSHRDGELAKASLDQFAVMFAALTR"
gene            10393. .10857
                /gene="asm28"
CDS             10393. .10857
                /gene="asm28"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAM54106.1"
                /db_xref="GI:21449374"
                /translation="MTDTTTRHRTPVQNAALAVGAVFLLVGIAGFVPGLTTNYDGLEF
                AGHHSDAALLGVFKVSALHNLVHLAFGAAGIAAGLSKAPRGARAFLIGGGAIYAVLWL
                YGVVVDEHSTANFVPLNAADNWLHLGLAVGMIGLGAALGRNAAHYRAARSDD"
gene            complement(10912. .11937)
                /gene="asm27"
CDS             complement(10912. .11937)
                /gene="asm27"
                /codon_start=1
```

```
                     /transl_table=11
                     /product="glucose 1-dehydrogenase"
                     /protein_id="AAM51105.1"
                     /db_xref="GI:21443373"
                     /translation="MPRRLAEQVVVVVGASSGIGRVTALRFAAAGARVVCAARNTRAL
                     DGLVEEVRGAGGRAVAVPADIADEAAVRAVADVAVERFGRVDTWVNAAGIGVYGRVED
                     TPAGEFDRVMRVNYLGHVHGAKAALPALRRAGGGVLIGVASVLGLRSAPLQAPYAASK

  Query Match             64.8%;  Score 65.4;  DB 1;  Length 82746;
  Best Local Similarity   78.8%;  Pred  No. 0.00031;
  Matches   78;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCCAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||| || || || |||||||| |||||| ||||||| |||||||| ||||||   ||
Db   71233 GCGCGGACTGTCCCCGGACGGGCGCTGCCAGGCGTTCTCCGCCGACGCCGACGGCACGGC 71174

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           | |||| |||||   || ||||||| |||| |||||||
Db   71173 GTGGGGTGAGGGCGTCGGCGTGGTGCTGCTGGAGCGGCT 71135


RESULT 15
SAU421825
LOCUS       SAU421825              66808 bp    DNA     linear   BCT 16-MAR-2002
DEFINITION  Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
            ORF1, stiA gene, stiB gene, stiC gene, stiD gene, stiE gene, stiF
            gene, stiG gene, stiH gene, stiJ gene, stiK gene, stiL gene and
            ORF9.
ACCESSION   AJ421825
VERSION     AJ421825.1  GI:19572309
KEYWORDS    acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
            kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
            ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
            channel beta chain; ribosome binding factor; stiA gene; StiA
            protein; stiB gene; StiB protein; stiC gene; StiC protein; stiD
            gene; StiD protein; stiE gene; StiE protein; stiF gene; StiF
            protein; stiG gene; StiG protein; stiH gene; StiH protein; stiJ
            gene; StiJ protein; stiK gene; stiL gene.
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1
  AUTHORS   Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
            Hofle,G. and Muller,R.
  TITLE     The biosynthesis of the aromatic myxobacterial electron transport
            inhibitor stigmatellin is directed by a novel type of modular
            polyketide synthase
  JOURNAL   Online Publication
  REMARK    J. Biol. Chem., 10.1074/jbc.M111738200
REFERENCE   2  (bases 1 to 66808)
  AUTHORS   Muller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
            Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .66808
                     /organism="Stigmatella aurantiaca"
                     /strain="Sg a15"
                     /db_xref="taxon:41"
     CDS             complement(4978. .6031)
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450 monooxygenase"
                     /protein_id="CAD19077.1"
                     /db_xref="GI:19572310"
                     /translation="MLRSVSASRPPSPANPVAAFDLARLDDAFYADPFPLYRAMRERD
                     PVHRMPDGSLFLTRWADLDCIYRDTRTFSSDKRAEFGAKYGDAPLFEHHTTSLVFNDP
                     PLHTRVRRLIVGALTPRALSTMEPGLRTLVDRLLDGLAVKGAADLIEDFAAAIPIEVI
                     GNLLDIPTEERGPLRGWSLAILGALEPRLTAEQEACGNEAVTEFLDYLRILVAQRRAR
                     PGDPATDVLTRLIQGESDGERLTETELLHQCVFLLNAGHETTTNLIGNALELLARFPD
                     ERARLLRAPALIPTAVEEVLRYESSNQLGNRRVAEDTEIGGVAVPTGTFLTLCIGAAN
                     RDPARFEDPEHFDVGRQPNRHLAFAGGAHTCAGMNLARIEARIALAAFLARFPDYALT
                     APPVRARRARFRGFTAMPTRLGSLR"
     CDS             6369. .7730
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAD19078.1"
                     /db_xref="GI:19572311"
                     /translation="MEELISEALPDGSRVGPWRIEGRAGYGTYGAVYRAHRVGRICPR
                     SVALKLARYPNDLRFEREAELLERVRHANVPRLLGHGTWRAGPQGEAHPYVVMQWVEG
                     VRLYQWNKERLRTAREVLRVLAQVARALEATHMRGGLHRDVKGENVLVSPEGHASLMD
                     FGCGTWEGAAPLTGGILAPGTKAYRSPEALRFHWNHRNEADAHYQATPADDVYALGVT
                     AYRLCTGIYPPPATEPSGVGEDERETQEVKKPPSQVRPLVPALEQLILRMLRDNPKER
                     GRAGELAAEMEALAERAGAEADEPMNPNHSSVPSEITGRPVSVQREWGGWPVALVLVA
                     VGLLLWGSGSLNTGRRVAPASRMTEGGTGGVADAAVEALPVSSAAREPEPSGLNLDML
                     KEPLPGQRRPPCPRFQISIQGGCWLEVPQSSPPCGESYYEWKGACYSPVAAPPRPRTS
                     DPQ"
     CDS             complement(7849. .8577)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl transferase"
                     /protein_id="CAD19079.1"
                     /db_xref="GI:19572312"
                     /translation="MSQNIYDDPSFFESYSQLARSLKGLAGAPEWPALRALLPDVKDL
                     RVVDLGCGFGWFCRWARKNDAAHVLGLDLSEKMLTRAKAETSGDAITYQQADLERLVL
                     PEGAFDLAFSSLAIHYVEELEPLLATVHRGLTPGGWFVFSIEHPIFMASRQPEWMTGP
                     EGHRTWPVDSYQLEGKRTTHWLAEEVVKYHRTLGTTLNALLEQGFTLRHIEEWGPTDA
                     QLSAHPELAEERDRPMFLLVSAQR"
     CDS             complement(8622. .8879)
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA binding protein"
                     /protein_id="CAD19080.1"
                     /db_xref="GI:19572313"
                     /translation="MALEDDFRSAQERVKALKTRPSNDRLLELYSLFKQATEGDVQGK
                     RPGILDVKERAKYDAWVGRKGLGRESAMQQYVALVERLLRG"
     CDS             8961. .10070
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="potassium channel beta chain"
                     /protein_id="CAD19081.1"
                     /db_xref="GI:19572314"
                     /translation="MVYLLCRSRSQLRWILNGSAHRVRFAWGGHARPRRMHFRHLGRS
                     GLMVSEISYGNWLTHGSQVEEAAAVACVRAALDEGITTFDTADVYAGTRAEEVLGRAL
                     QGQRREGLEIFTKVYWPTGPGPNDRGLSRKHILESIHGSLRRLKTEYVDLYQAHRYDT
                     ETPLEETMQAFSDVVRQGKAHYIGVSEWTADQLREGVKLSRQMGFQLVSSQPQYSMLW
                     RVIEAEVVPASQELGVGQVVYSPIAQGVLTGKYRPGQKPPEGSRATDEKGGASFIKRF
                     LNDDVLTRVQQLTPVAQEAGLSLAQLAVAWVLQNPNVSSAIIGASRPEQVKENVKAAG
                     VKLSEEAMRHIDTLLGPIVERNPKKTLENSPRKRP"
     CDS             10428. .10862
                     /note="ORF3"
                     /codon_start=1
                     /transl_table=11
                     /product="ribosome binding factor"
                     /protein_id="CAD19082.1"
                     /db_xref="GI:19572315"
                     /translation="MTSFKQRRRRAPRRGTESSLSSFEAQSLPAHHLRVQSTLFQEVA
                     RLFRGELSDPLLEGVSVTSFELTPDGRNARIGFTLPPEYAEGGSAPIEAALVRASGFI
                     RAQLALGLNLKRVPHLRFVCVGVAPPGFSIPEDGSAEEGGES"
     CDS             10859. .12016
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD19083.1"
                     /db_xref="GI:19572316"
                     /translation="MRPRRIEGAPGAVPILAWARAVPAGAVKQLQLLARQPYLAEHVA
                     AMPDVHVAEGVAVGTVFATERHIVPGALGGDLGCGVSAVRFDFPAMSLGRVELTRLLA
                     AFARAVPVGDAVHRGRGLSLPPGLDIGRLSTHRLQREWERLVPKHLGTLGGGNHFLEL
                     DRDAEGALWLLLHTGSRGVGAAIAAHHLRVAQVSGEGSLPGLRTDTSEGAACLADTEL
```

```
                ACQFARANRDTLLQRAVEVLAGAIGQGPVLDSRVDVHHNHVAAESHFGRTLLVHRKGA
                VGLVAGQQGLIPGSMGTASYLVAGKGEPQAFCSCSHGAGRVMTRREARERIRPAALAH
                ALRRVVFDEGRVTALVEEAPTAYRDITEVLEDEAELVTPLRRLTPIAVLKG"
CDS             complement(12082. .13974)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="cellulase"
                /protein_id="CAD19084.1"
                /db_xref="GI:19572317"
                /translation="MKNSLLLLMLAVLPLYGTAALARPGMANAEPGHFEGPENAGLTA
                AQCNPRVPLSTRGRYIVDRCGERFKLKSVNWFGASDQLEVVGGLDKQKLSDLVAGMKA
                MGFNSVRLPFSNNMLHPDDNKPAAQRCLEKYSVTCIDSTKNPELLNKTALEVYDAVVA
                ELTRQGLVVILNNHTTKSMWCCGWDGNGFWDSSQALQRWQDDWVMLAARYQGNKWVAG
                ADLRNEVRPDGLDSPNWGMRNQHDWHMAAQTMGNLLLRTNPDLLIVVEAVNWWGLLDG
                SRPQLKPVRQRPVALLRGDKLVYAVHNYGYTGPNQSGGSLGSGPKYSDMDKPTLYGTL
                DQEWGFALAANQAYTAPVWMSEFGVGYNEQAANSRAWFSNLADYLIEKDVDWAYWAIN
                AAKLQNSVQGEDETYGLWSYPDWNGVRSGDWRLGTGPLGRLLGADGRTGAVDATRFLP
                MTVLLKDGSSTYYVDNNSLDSDWHSGKAKISCPRGYRMVGVSANFTILCTNAGAVPAQ
                QGTGSYQTITQEVSPLRYPGDWANQSIKFECPTGFYAVGISTANPFFLELAGLLCEQN
                TGGIPLNSKSAKNFWGGDQRASSSGGDWSVSQYKGQCADNQYLIGLAHRRSGLADYPS
                VVLCSN"
gene            14345. .21466
                /gene="stiA"
CDS             14345. .21466
                /gene="stiA"
                /function="PKS"
                /codon_start=1
                /transl_table=11
                /product="StiA protein"
                /protein_id="CAD19085.1"
                /db_xref="GI:19572318"
                /translation="MTQLASRLGVGSESIDPRERFSRYGIDSRDTKSLLAELGALLKR
                PLSPTLAWTYPTVEALSRHLAGEADVSGGERADTLGVDQDEPIAIIGMAGRFPKAPDV
                ESFWSLLRDGVDAITEVPPERWDLDAFYSVGLKSPGKMNTRWGGFLDQVDQFDPLFFG
                ISPREAIQMDPQQRLALELSWEALESAGVAVDSLMDSRTGVFMGAMWSDYAKLTRGAL
                GLIGPHTATGQDSSIIAARVSYFLGLRGPSLTVNTACSSALVAVHLACQSLRNGEAAM
                ALAGGVNLLLSPESSVAMSQLGALSPDGRCKAFDARANGYVRGEGGGVVVLTPLSRAI
                ANGYPVLCVIRGSAVNNDGFSNGLTAPNPRAQEEVLRDAYARAGVSPDKVHYVETHGT
                GTVLGDPIEANALSAVLGVNRPAERPLILGSVKTNIGHLEAASGIAGLMKVVLSIQHR
                AIPPNLHFEQPNPYIRFEELRLLVPKSLHPWPNEGESALAGISSFGFGGTNCHLVLEA
                PSPGSAHIAPLAAETPEELSLLARQARDAVSVVGSQVELADLCFSVAARCREGSQRLA
                LIARSPAELAAQVDGFLAGEKRPGLLRGEVTSGGSQKPVFVFSGTGSQWPGMGRELLR
                TEPVFRGAIERCDRALRGLSGWSMLEELSAGEASSRLKECSGLQTAIFAIQVGLAALW
                RSWGIEPAGVVGHSIGEVAAAHVAGILNLEDAARVIHHLSRLMGRLSGQGAMIVVDLS
                AEEANRLLREWERHVVVAACNAPASTVLSGAREVIEELGAELERRGIAAHTVRIDTPS
                HSPLMDPFMGELAEVLAGLSPRVAAVPMVSTVTGAFLKAEECTAAYWVRNLREPVLFS
                QALTKLFEQDLRLFLELSPHPILARSVEQCLKGEGRQGVVLASMRRGENEGACMRESL
                GRLFTLGTSVAWQALYPAEAKRVELPMGLGTGSLDAELRGSARLLPLSAHTAEALEDA
                ARALQASIAKHPDVRLDDLCYSASVGKSVHAHRLAFVTDSKRALVEHLDDFLKGESRP
                GRIAGVCATGEPKKVVFVFPGQGSQWQGMGRQLLRQEPVFREAIEECDAAIRSCAGWS
                LIDALQADEPHSRLGSIEVIQPTLFAIEVALAALWRSWGVVPDAVVGHSMGETAAACV
                AGALRLTDAARVICERSQLLRRVAGQGGMALVELSLEQAREMLSRKDGRVSIGASNSP
                RSTVLSGDAIALEEILRELEAQGTFCRRVKVDVASHSPQVDPLKDDLLRVLSTIAPRA
                ATVPLYSTVTGEPIEGARMDASYWMRNLREPVLFGQAIERLAGDGHGVFVELSPHPIL
                LPSIDQTLRHKSHKGVVLPSLRREEPERAVLLESLGTLFTGGGRVDFRGLHPLGGHRV
                RLPHLPWQRQRFWLDMDVAAAGAASVGAAWHPGGGSEHPLLGHRLPSLAHAPACHVWE
                NKLSRSAVSYLDDHRVGGTVVLPGTSYIEMALAAAAKGSGKKTVSLRDVELREKMSFL
                ETASRTVQLSLIQEEAGHSSFRLHGRSTPDEPWTLHGGGTVHVEEGDGPPSADPVLWA
                AVETVRLQGKEVSASDAYEAMHAQGNEYGPLFRGIEQMWVRDAELVARLRVPEPLEQE
                LPEYRFHPAALDLGLQALIVLAAEHGCEGLALPTRLESIRAHASPGARMWCHVRLNPV
                REAASFSGEVLFLDESGQRLAELQGARFQILGEGSRPHDGGRLVDWLHEVEWQPAPVP
                VPTAEPSSQGRWVLFADGTGVAEGLAAQIAARGEPCIRVLPGMAFAEAEDGHFVVDPE
                QPEDVRKLFGALSRTSTIPLRGIVHLWSLDAPSPEEGGVDALSRARRLGSGSVLSLIQ
                GLNAAELAQKPRLWLVTRGVQAVGSAGPGAGAVAQSPLWGLGQVVAYEHAELRCTRVD
                LAPVSVPGEEQTLLQELLAPAGDDQVALRGRERHVPRLTRCRTDEVAGRSTDLRAGAI
                RSDGCYLITGGLGGLGLTIAKWLVARGARHLVLVGRRGASDAAEIALGQMRGAGAECL
                VMKADVANTSEVEELVGTIQRKMPPLRGIFHAAGLLDDGTVLQLTQARLRSVMAPKVE
                GAWNLHRATLGASLDFFVLFSSAASLVGIPGQGNYAAGNAFLDALAGLRKARGLACQS
                IQWGPWSEIGLAAEQDNRGERLALRGMGSLNPREGLEALEALLSPGEPLRERARIGVL
                RLDARRWCKSYPAIRSSPLFSVLLAGERNAKQVLERSGLRARLNEEPARTRRAALESF
                LREEAAQVLRLPPARIDLSRSLQTLGMDSLLSLELRNRLEANLELTLPATVLWNYPTL
                SALAPFLAERMGIQLNGVEEPPQAVESPSVSEEDGAAMLKMLEQFQELSNEDLLGMLS
                DDKEDEVSS"
    gene        21463. .26220
                /gene="stiB"
    CDS         21463. .26220

  Query Match          63.8%;  Score 64.4;  DB 1;  Length 66808;
  Best Local Similarity  78.6%;  Pred. No. 0.00051;
  Matches   77;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           || || ||  |||| || ||||||||||||||||||   ||||  ||| ||||||| | |
Db   50961 CGGGCCCTCGCGCCTGATGGGCGGTGCAAGACGTTCGACGCCGCGGCGAACGGCTATGTG 51020

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           ||| | || |||||||   ||||||||||||||| ||
Db   51021 CGGAGTGAAGGGTGCGGAATGGTGGTGCTCAAGCGCCT 51058


Search completed: November  5, 2002, 12:26:41
Job time : 617.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_25850_25950
Perfect score: 101
Sequence: 1 gcgcgcgctttcgcccgacg.....ggtggtgctcaagcggctgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
2	101	100.0	71989	21	AAA29349	Sorangium cellulos
3	65.4	64.8	33529	23	AAS17367	DNA sequence of S.
4	62.6	62.0	764	22	AAF89974	Partial nucleotide
5	62	61.4	6564	22	AAF31637	Mycobacterium tube
c 6	61.4	60.8	4403765	22	AAI99683	Mycobacterium tube
c 7	61.4	60.8	4411529	22	AAI99682	Mycobacterium tube
8	61.2	60.6	758	22	AAF89970	Partial nucleotide
9	61.2	60.6	5088	22	AAF90038	Nucleotide sequenc
10	61.2	60.6	34071	22	AAF90033	Nucleotide sequenc
c 11	61.2	60.6	42717	22	AAF90032	Nucleotide sequenc
c 12	60.8	60.2	333	21	ABQ62819	Mycobacterium tube
13	60.8	60.2	5481	22	AAF31638	Mycobacterium tube
14	60.8	60.2	5485	22	AAH52061	Mycobacterium tube
15	60.2	59.6	109519	22	AAS08693	Micromonospora DNA
16	60	59.4	29879	14	AAQ46806	eryA region of S.
17	59.6	59.0	671	22	AAF89977	Partial nucleotide
18	59.6	59.0	49377	19	AAV05287	The soraphen biosy
c 19	57.8	57.2	507	21	ABQ63181	Mycobacterium tube
c 20	57.4	56.8	14071	23	ABL50560	Micromonospora car
21	57.4	56.8	28598	17	AAT06769	Sorangium cellulos
22	57.4	56.8	28958	18	AAT89956	Sorangium cellulos
23	57.4	56.8	28958	21	AAA75299	DNA sequence of So
24	56.8	56.2	671	22	AAF89969	Partial nucleotide
25	56.8	56.2	4306	22	AAF90039	Nucleotide sequenc
26	56.8	56.2	47981	22	AAF30757	Micromonospora meg
27	56.8	56.2	53789	19	AAV21187	Amycolatopsis medi
28	56	55.4	15872	18	AAT68715	Streptomyces venez
29	56	55.4	15872	21	AAZ87283	S. venezuelae vep
30	56	55.4	43280	18	AAT80413	Tylactone synthase
31	55.8	55.2	751	21	AAA58914	DNA encoding ketos
32	55.8	55.2	14775	22	AAF88338	S. spinosa DNA fra
33	55.8	55.2	20394	22	AAF24892	Pimaricin biosynth
34	55.8	55.2	50000	22	AAF88313	S. spinosa DNA fra
35	55.8	55.2	50000	22	AAF88316	S. spinosa DNA fra
36	55.8	55.2	80161	20	AAZ21501	DNA fragment of Sa
37	55.2	54.7	13987	18	AAT80415	Hybrid srmG/tylG O
38	54.8	54.3	671	22	AAF89966	Partial nucleotide
39	54.2	53.7	44377	18	AAT78508	Platenolide syntha
40	54.2	53.7	44377	18	AAT80414	Platenolide syntha
41	53.6	53.1	672	22	AAF89964	Partial nucleotide
42	53.6	53.1	65140	22	AAD17184	Streptomyces nours
43	53.6	53.1	125401	22	AAD17186	Streptomyces nours
44	53.4	52.9	5292	22	AAF90036	Nucleotide sequenc
45	52.6	52.1	668	22	AAF89976	Partial nucleotide

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;    99WO-EP04171.
XX
PR   18-JUN-1998;    98US-0099504.
PR   24-SEP-1998;    98US-0101631.
PR   05-FEB-1999;    99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes ER5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred  No. 5.6e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25850 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 25909

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25910 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 25950


RESULT 3
AAS17367
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   12-MAR-2002  (first entry)
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
XX
OS   Sorangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT   -
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the DNA
CC   sequence of cosmid pKOS28-26 which encodes one or more domains of
CC   S. cellulosum PKS.
CC   Note: The present sequence is said to encode the functional domains
CC   of S. cellulosum PKS which correspond to domains or domain subsets of
CC   the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

  Query Match             64.8%;  Score 65.4;  DB 23;  Length 33529;
  Best Local Similarity   78.8%;  Pred. No. 2.1e-07;
  Matches   78;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||| ||| | |||||||||||| ||||||||||||||||||   |  || ||||||||||
Db   10351 GCGTGCGGTCTCGCCCGACGGACGGTGCAAGACGTTCTCGCAGGCGGCCGACGGCTACGG 10410

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             || ||||||||   ||  ||||||||||| | ||| ||
Db   10411 CCGCGGCGAGGGCGTCGGGGTGGTGGTGCTGATGCGTCT 10449


RESULT 4
AAF89974
ID   AAF89974 standard; DNA; 764 BP.
XX
AC   AAF89974;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
```

```
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83966.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 224; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 764 BP; 114 A; 272 C; 268 G; 110 T; 0 other;

  Query Match             62.0%;  Score 62.6;  DB 22;  Length 764;
  Best Local Similarity   79.6%;  Pred. No. 1.2e-06;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       7 GCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGG 66
           |||  |||| |||||||| ||||||||||||   ||||  ||||||||||||  || ||
Db     480 GCTCGCGCCTGACGGGCGCTGCAAGACGTTCGACGCCGCGGCGGACGGCTACGTCCGCGG 539

Qy      67 CGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           ||| |||  ||| ||| | ||||| ||||||||
Db     540 CGAAGGGGCCGCGGTGATCGTGCTGAAGCGGCT 572


RESULT 5
AAF31637
ID   AAF31637 standard; DNA; 6564 BP.
XX
AC   AAF31637;
XX
DT   09-APR-2001  (first entry)
XX
DE   Mycobacterium tuberculosis ppsC gene.
XX
KW   Mycobacterium tuberculosis; attenuated microorganism;
KW   signature tagged transposon mutant; mutant library;
KW   mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW   vaccine; ppsC; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200102555-A1.
XX
PD   11-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-IB00950.
XX
PR   06-JUL-1999;   99US-0142982.
PR   08-JUL-1999;   99US-0142833.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Guilhot C,  Camacho L;
XX
DR   WPI; 2001-091804/10.
DR   P-PSDB; AAB66467.
XX
PT   Screening a mutant library for mutants unable to grow under specific
PT   conditions and for identifying loci involved in pathogenicity,
PT   comprises using signature tagged transposon mutagenesis  -
XX
PS   Example 8; Page 122-124; 159pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   screening a library of mutants. The method comprises constructing a
CC   library with insertions in genes and/or regulatory regions of the
CC   organisms of interest, where the insertion contains a tag and/or a
CC   transposon associated with a tag. The mutants are identified by
CC   hybridisation of the tags to known sequences. The method is useful for
CC   treating an individual suffering from a mycobacterial infection,
CC   suspected of being infected with a Mycobacterium, or having been
CC   exposed to an infectious Mycobacterium. It is also useful for
CC   identifying and isolating mutants of actinomycetales and for identifying
CC   compounds that have antibiotic activity. The method is used to identify
CC   mutants of microorganisms, preferably an actinomycetales, such as
CC   M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC   M. paratuberculosis, that is unable to grow under specific conditions.
CC   It is especially useful for identifying loci involved in pathogenicity.
CC   It is useful in constructing vaccines. The method can be used to screen
CC   multiple libraries concurrently. It can screen libraries of different
CC   organisms or different strains of the same organism.
XX
SQ   Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 other;

  Query Match             61.4%;  Score 62;  DB 22;  Length 6564;
  Best Local Similarity   78.7%;  Pred. No. 1.5e-06;
  Matches   74;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy       7 GCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGG 66
           ||| || || || ||||||||||||| |||   ||   ||| || || |||| ||| ||
Db     762 GCTGTCACCGGAGGGGCGGTGCAAGACCTTCGATGCGTCCGCCGATGGATACGTGCGCGG 821

Qy      67 CGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           ||||||   ||| ||||||||||||||||||||
Db     822 CGAGGGTGCCGCGGTGGTGGTGCTCAAGCGGCTG 855


RESULT 6
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
```

```
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             60.8%;  Score 61.4;  DB 22;  Length 4403765;
  Best Local Similarity   77.9%;  Pred. No. 1.4e-06;
  Matches   74;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           | |||  |||||||||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4252406 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4252347

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             ||| ||  ||| | || ||||||||||||||| |
Db 4252346 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGGGT 4252312


RESULT 7
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             60.8%;  Score 61.4;  DB 22;  Length 4411529;
  Best Local Similarity   77.9%;  Pred. No. 1.4e-06;
  Matches   74;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           | |||  |||||||||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4260150 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4260091

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             ||| ||  ||| | || ||||||||||||||| |
Db 4260090 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGGGT 4260056


RESULT 8
AAF89970
ID   AAF89970 standard; DNA; 758 BP.
XX
AC   AAF89970;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83962.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 222; 356pp; French.
XX
```

```
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 758 BP; 121 A; 250 C; 247 G; 140 T; 0 other;

  Query Match             60.6%;  Score 61.2;  DB 22;  Length 758;
  Best Local Similarity   76.5%;  Pred. No. 2.7e-06;
  Matches   75;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           |||  | |  |||||||||| || ||||||||||||   ||||  || ||||| || | |
Db     469 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 528

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || |||||||| |||| | | ||| |||| ||||||||
Db     529 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGCT 566


RESULT 9
AAF90038
ID   AAF90038 standard; DNA; 5088 BP.
XX
AC   AAF90038;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5088
FT                   /*tag= a
FT                   /product= "type I polyketide synthase"
FT                   /transl_except= "(1..3, aa: Met)"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83975.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 309-311; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 5088 BP; 838 A; 1622 C; 1679 G; 949 T; 0 other;

  Query Match             60.6%;  Score 61.2;  DB 22;  Length 5088;
  Best Local Similarity   76.5%;  Pred. No. 2.4e-06;
  Matches   75;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           |||  | |  |||||||||| || ||||||||||||   ||||  || ||||| || | |
Db     739 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 798

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || |||||||| |||| | | ||| |||| ||||||||
Db     799 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGCT 836


RESULT 10
AAF90033
ID   AAF90033 standard; DNA; 34071 BP.
XX
AC   AAF90033;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Example 14; Page 289-300; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
```

```
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 coding strand), which encodes different
CC   polyketide synthases.
XX
SQ   Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

  Query Match             60.6%;  Score 61.2;  DB 22;  Length 34071;
  Best Local Similarity   76.5%;  Pred. No. 2.1e-06;
  Matches   75;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           |||  | |  |||||||||| || ||||||||||||   ||||  || ||||| || | |
Db   25420 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 25479

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || |||||||| |||| | | ||| |||| ||||||||
Db   25480 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGCT 25517


RESULT 11
AAF90032/c
ID   AAF90032 standard; DNA; 42717 BP.
XX
AC   AAF90032;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Example 14; Page 274-288; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC   encodes different polyketide synthases.
XX
SQ   Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

  Query Match             60.6%;  Score 61.2;  DB 22;  Length 42717;
  Best Local Similarity   76.5%;  Pred. No. 2.1e-06;
  Matches   75;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           |||  | |  |||||||||| || ||||||||||||   ||||  || ||||| || | |
Db   8700 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 8641

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || |||||||| |||| | | ||| |||| ||||||||
Db   8640 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGCT 8603


RESULT 12
ABQ62819/c
ID   ABQ62819 standard; DNA; 333 BP.
XX
AC   ABQ62819;
XX
DT   16-AUG-2002  (first entry)
XX
DE   Mycobacterium tuberculosis BAC vector clone Rv272T7.
XX
KW   Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
KW   detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW   gene; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9954487-A2.
XX
PD   28-OCT-1999.
XX
PF   16-APR-1999;   99WO-IB00740.
XX
PR   16-APR-1998;   98US-0060756.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Cole S,  Buchrieser-Brosch R,  Gordon S,  Billault A;
XX
DR   WPI; 2000-013262/01.
XX
PT   Isolation of polynucleotides from mycobacterial genomes, useful for
PT   detection of Mycobacteria and for combating tuberculosis -
XX
PS   Claim 23; Page 77; 161pp; English.
XX
CC   The present invention describes a method for isolating a polynucleotide
CC   of interest that is present or is expressed in a genome of a first
CC   mycobacterium strain and that is absent or altered in a genome of a
CC   second mycobacterium strain, which is different from the first strain
CC   using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
CC   vectors, which are preferably immobilised, can be used to detect
CC   mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
CC   samples. The polynucleotides identified are useful as probes or primers
CC   for detecting a given mycobacterium of interest. By aligning the
CC   polynucleotides contained in the recombinant BAC vectors it is possible
CC   to physically map a polynucleotide of mycobacterial origin in a
CC   biological sample. The methods and vectors from the present invention
CC   are useful in providing information for combating tuberculosis. It is
CC   possible to compare genomes between different strains or species and
CC   their non-pathogenic strains or species counterparts. ABQ62492 to
CC   ABQ63228 and ABB81227 to ABB81230 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 333 BP; 66 A; 117 C; 102 G; 48 T; 0 other;

  Query Match             60.2%;  Score 60.8;  DB 21;  Length 333;
  Best Local Similarity   77.1%;  Pred. No. 3.6e-06;
```

```
  Matches   74;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCCCGG 64
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     182 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTCCGC 123

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           |||||||| |||| ||||||||||||||||||| ||
Db     122 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGGTTG 87


RESULT 13
AAF31638
ID   AAF31638 standard; DNA; 5481 BP.
XX
AC   AAF31638;
XX
DT   09-APR-2001  (first entry)
XX
DE   Mycobacterium tuberculosis ppsD gene.
XX
KW   Mycobacterium tuberculosis; attenuated microorganism;
KW   signature tagged transposon mutant; mutant library;
KW   mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW   vaccine; ppsD; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200102555-A1.
XX
PD   11-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-IB00950.
XX
PR   06-JUL-1999;   99US-0142982.
PR   08-JUL-1999;   99US-0142833.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Guilhot C,  Camacho L;
XX
DR   WPI; 2001-091804/10.
DR   P-PSDB; AAB66468.
XX
PT   Screening a mutant library for mutants unable to grow under specific
PT   conditions and for identifying loci involved in pathogenicity,
PT   comprises using signature tagged transposon mutagenesis  -
XX
PS   Example 8; Page 131-132; 159pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   screening a library of mutants. The method comprises constructing a
CC   library with insertions in genes and/or regulatory regions of the
CC   organisms of interest, where the insertion contains a tag and/or a
CC   transposon associated with a tag. The mutants are identified by
CC   hybridisation of the tags to known sequences. The method is useful for
CC   treating an individual suffering from a mycobacterial infection,
CC   suspected of being infected with a Mycobacterium, or having been
CC   exposed to an infectious Mycobacterium. It is also useful for
CC   identifying and isolating mutants of actinomycetales and for identifying
CC   compounds that have antibiotic activity. The method is used to identify
CC   mutants of microorganisms, preferably an actinomycetales, such as
CC   M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC   M. paratuberculosis, that is unable to grow under specific conditions.
CC   It is especially useful for identifying loci involved in pathogenicity.
CC   It is useful in constructing vaccines. The method can be used to screen
CC   multiple libraries concurrently. It can screen libraries of different
CC   organisms or different strains of the same organism.
XX
SQ   Sequence 5481 BP; 894 A; 1745 C; 1925 G; 917 T; 0 other;

  Query Match             60.2%;  Score 60.8;  DB 22;  Length 5481;
  Best Local Similarity   77.1%;  Pred. No. 3e-06;
  Matches   74;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     751 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 810

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           |||||||| |||| ||||||||||||||||||| ||
Db     811 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGGTTG 846


RESULT 14
AAH52061
ID   AAH52061 standard; DNA; 5485 BP.
XX
AC   AAH52061;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Mycobacterium tuberculosis potential drug target gene SEQ ID 115.
XX
KW   Drug target; growth; organism viability; characterisation; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200135317-A1.
XX
PD   17-MAY-2001.
XX
PF   13-NOV-2000; 2000WO-US31152.
XX
PR   12-NOV-1999;   99US-0165086.
PR   12-NOV-1999;   99US-0165124.
PR   01-FEB-2000; 2000US-0179531.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Eisenberg D,  Rotstein SH,  Marcotte EM;
XX
DR   WPI; 2001-329193/34.
DR   P-PSDB; AAG81210.
XX
PT   Identifying nucleotide or polypeptide sequence for use as drug target,
PT   involves providing algorithm that analyzes a functional relationship
PT   between nucleotide or polypeptide sequences, and comparing the
PT   sequences  -
XX
PS   Disclosure; Page 128-130; 207pp; English.
XX
CC   This invention relates to a method for identifying a nucleotide or
CC   polypeptide sequence that may be a drug target, or essential for growth
CC   or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC   represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC   tuberculosis proteins which are potential drug targets. The DNA and
CC   protein sequences are used to illustrate the method of the invention. The
CC   method involves providing an unknown nucleotide or polypeptide sequences,
CC   and comparing it to a number of sequences along with at least one
CC   algorithm capable of analysing a functional relationship between
CC   nucleotide and polypeptide sequences. The method is useful for
CC   characterising the function of nucleic acids and polypeptides that may be
CC   useful as a target for a drug or essential for the growth or viability of
CC   an organism.
XX
SQ   Sequence 5485 BP; 896 A; 1745 C; 1926 G; 918 T; 0 other;

  Query Match             60.2%;  Score 60.8;  DB 22;  Length 5485;
  Best Local Similarity   77.1%;  Pred. No. 3e-06;
  Matches   74;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     751 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 810
```

```
Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           |||||||| |||| ||||||||||||||||||| ||
Db     811 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGGTTG 846

RESULT 15
AAS08693
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW   Everninomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (132..1382)
FT                   /*tag= a
FT                   /product= "EvdA"
FT   RBS             complement (1389..1394)
FT                   /*tag= b
FT   CDS             complement (1490..2611)
FT                   /*tag= c
FT                   /product= "EvdB"
FT   RBS             complement (2618..2622)
FT                   /*tag= d
FT   CDS             complement (2622..3860)
FT                   /*tag= e
FT                   /product= "EvdC"
FT   RBS             complement (3867..3870)
FT                   /*tag= f
FT   CDS             4143..5312
FT                   /*tag= g
FT                   /product= "EvdD"
FT   RBS             4134..4138
FT                   /*tag= h
FT   CDS             5309..6235
FT                   /*tag= i
FT                   /product= "EvdE"
FT   CDS             6232..7275
FT                   /*tag= j
FT                   /product= "EvdF"
FT   RBS             6226..6229
FT                   /*tag= k
FT   CDS             7272..8327
FT                   /*tag= l
FT                   /product= "EvdG"
FT   CDS             8342..9364
FT                   /*tag= m
FT                   /product= "EvdH"
FT   RBS             8333..8336
FT                   /*tag= n
FT   CDS             complement (9463..10224)
FT                   /*tag= o
FT                   /product= "EvdI"
FT   RBS             complement (10232..10235)
FT                   /*tag= p
FT   CDS             10424..11176
FT                   /*tag= q
FT                   /product= "EvdJ"
FT   CDS             12027..12455
FT                   /*tag= r
FT                   /product= "EvdK"
FT                   /partial
FT                   /note= "No start codon"
FT   CDS             complement (12108..13022)
FT                   /*tag= s
FT                   /product= "EvdL"
FT   RBS             complement (13027..13030)
FT                   /*tag= t
FT   CDS             complement (14410..15363)
FT                   /*tag= u
FT                   /product= "EvrA"
FT   RBS             complement (15369..15373)
FT                   /*tag= v
FT   CDS             complement (15380..16414)
FT                   /*tag= w
FT                   /product= "EvrB"
FT   CDS             complement 16419..17873
FT                   /*tag= x
FT                   /product= "EvrC"
FT   CDS             complement (17870..18934)
FT                   /*tag= y
FT                   /product= "EvrD"
FT   CDS             19374..20906
FT                   /*tag= z
FT                   /product= "EvrE"
FT   CDS             21064..22542
FT                   /*tag= aa
FT                   /product= "EvrF"
FT   RBS             21056..22542
FT                   /*tag= ab
FT   CDS             22748..24172
FT                   /*tag= ac
FT                   /product= "EvrG"
FT   RBS             22736..22740
FT                   /*tag= ad
FT   CDS             complement (24177..25223)
FT                   /*tag= ae
FT                   /product= "EvrH"
FT   RBS             complement (25230..25233)
FT                   /*tag= af
FT   CDS             25550..26626
FT                   /*tag= ag
FT                   /product= "EvrI"
FT   CDS             26685..30479
FT                   /*tag= ah
FT                   /product= "EvrJ"
FT   RBS             26672..26676
FT                   /*tag= ai
FT   CDS             complement (30557..31876)
FT                   /*tag= aj
FT                   /product= "EvrK"
FT   RBS             complement (31885..31888)
FT                   /*tag= ak
FT   CDS             complement (31941..32882)
FT                   /*tag= al
FT                   /product= "EvrL"
FT   CDS             complement (33167..34405)
FT                   /*tag= am
FT                   /product= "EvrM"
FT   RBS             complement (34414..34418
FT                   /*tag= an
FT   CDS             complement (34449..35210)
FT                   /*tag= ao
FT                   /product= "EvrN"
FT   RBS             complement (35219..35221)
FT                   /*tag= ap
FT   CDS             complement (35294..36238)
FT                   /*tag= aq
FT                   /product= "EvrO"
FT   CDS             complement (36235..36963)
FT                   /*tag= ar
FT                   /product= "EvrP"
FT   CDS             complement (36998..38026)
FT                   /*tag= as
FT                   /product= "EvrQ"
FT   CDS             complement (38072..38566)
FT                   /*tag= at
FT                   /product= "EvrR"
FT   CDS             complement (38892..40163)
```

```
FT                   /*tag= au
FT                   /product= "EvrS"
FT   CDS             complement (40216..40890)
FT                   /*tag= av
FT                   /product= "EvrT"
FT   RBS             complement (40899..40902)
FT                   /*tag= aw
FT   CDS             complement (40887..41576)
FT                   /*tag= ax
FT                   /product= "EvrU"
FT   CDS             complement (41679..42707)
FT                   /*tag= ay
FT                   /product= "EvrV"
FT   RBS             complement (42714..42717)
FT                   /*tag= az
FT   CDS             complement (42810..43799)
FT                   /*tag= ba
FT                   /product= "EvrW"
FT   RBS             complement (43807..43811)
FT                   /*tag= bb
FT   CDS             complement (43799..44866)
FT                   /*tag= bc
FT                   /product= "EvrX"
FT   CDS             complement (45014..45760)
FT                   /*tag= bd
FT                   /product= "EvrY"
FT   RBS             complement (45767..45770)
FT                   /*tag= be
FT   CDS             complement (45962..46714)
FT                   /*tag= bf
FT                   /product= "EvrZ"
FT   RBS             complement (45952..45956)
FT                   /*tag= bg
FT   CDS             complement (47156..49234)
FT                   /*tag= bh
FT                   /product= "EvsA"
FT   CDS             51627..52715
FT                   /*tag= bi
FT                   /product= "EvsB"
FT   RBS             51629..51622
FT                   /*tag= bj
FT   CDS             52889..53557
FT                   /*tag= bk
FT                   /product= "EvsC"
FT   CDS             53554..54207
FT                   /*tag= bl
FT                   /product= "EvbA"
FT   CDS             complement (54362..55117)
FT                   /*tag= bm
FT                   /product= "EvbB"
FT   RBS             complement (55125..55128)
FT                   /*tag= bn
FT   CDS             complement (55135..56094)
FT                   /*tag= bo
FT                   /product= "EvbC"
FT   RBS             complement (56100..56103)
FT                   /*tag= bp
FT   CDS             complement (56184..56813)
FT                   /*tag= bq
FT                   /product= "EvbC2"
FT   CDS             56961..58709

  Query Match             59.6%;  Score 60.2;  DB 22;  Length 109519;
  Best Local Similarity   76.3%;  Pred. No. 3.4e-06;
  Matches   74;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       3 GCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGC 62
           | || ||  | |||||||||||  ||||| | |||   |||  ||| ||||||||||  |
Db   27367 GTGCCCTGGCACCCGACGGGCGCAGCAAGTCCTTCGACGCCACCGCCGACGGCTACGGTC 27426

Qy      63 GGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           | |||||||||||||| |||  | |||||||||| |||
Db   27427 GTGGCGAGGGGTGCGGCGTCCTCGTGCTCAAGCTGCT 27463
```

```
Search completed: November  5, 2002, 13:36:41
Job time : 1694.22 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_25850_25950
Perfect score: 101
Sequence: 1 gcgcgcgctttcgcccgacg.....ggtggtgctcaagcggctgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
9	65.4	64.8	33529	4	US-09-144-085-3	Sequence 3, Appli
c 10	61.4	60.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 11	61.4	60.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 12	60.8	60.2	333	4	US-09-060-756-318	Sequence 318, App
13	60	59.4	11219	1	US-07-642-734C-1	Sequence 1, Appli
14	60	59.4	11219	3	US-08-439-009A-1	Sequence 1, Appli
15	59.6	59.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
c 16	57.8	57.2	507	4	US-09-060-756-680	Sequence 680, App
17	57.4	56.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
18	57.4	56.8	28958	1	US-08-456-837-6	Sequence 6, Appli
19	57.4	56.8	28958	1	US-08-457-342-6	Sequence 6, Appli
20	57.4	56.8	28958	1	US-08-457-646A-6	Sequence 6, Appli
21	57.4	56.8	28958	1	US-08-458-076A-6	Sequence 6, Appli
22	57.4	56.8	28958	1	US-08-764-233A-4	Sequence 4, Appli
23	57.4	56.8	28958	1	US-08-457-335A-6	Sequence 6, Appli
24	57.4	56.8	28958	1	US-08-729-214-6	Sequence 6, Appli
25	57.4	56.8	28958	3	US-09-028-934-6	Sequence 6, Appli
26	56	55.4	15872	4	US-09-105-537-1	Sequence 1, Appli
27	56	55.4	43280	2	US-08-804-227C-1	Sequence 1, Appli
28	55.8	55.2	751	3	US-09-010-809-4	Sequence 4, Appli
29	55.8	55.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
30	55.8	55.2	80161	4	US-09-370-700-1	Sequence 1, Appli
31	55.2	54.7	13987	2	US-08-804-227C-13	Sequence 13, Appl
32	54.2	53.7	20235	1	US-07-642-734C-3	Sequence 3, Appli
33	54.2	53.7	20235	3	US-08-439-009A-3	Sequence 3, Appli
34	54.2	53.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
35	54.2	53.7	44377	2	US-08-804-198-1	Sequence 1, Appli
36	49.8	49.3	50937	4	US-09-428-517-1	Sequence 1, Appli
37	47.8	47.3	423	4	US-09-144-085-5	Sequence 5, Appli
38	47.8	47.3	13842	4	US-09-105-537-30	Sequence 30, Appl
39	47.8	47.3	36778	4	US-09-105-537-5	Sequence 5, Appli
40	47.8	47.3	38506	3	US-09-320-878-19	Sequence 19, Appl
41	45.6	45.1	11220	4	US-09-105-537-32	Sequence 32, Appl
42	44.6	44.2	4041	4	US-09-105-537-36	Sequence 36, Appl
43	44	43.6	1462	4	US-09-434-288-4	Sequence 4, Appli
44	44	43.6	4689	4	US-09-105-537-34	Sequence 34, Appl
45	42.8	42.4	643	4	US-08-861-774E-53	Sequence 53, Appl

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 5.3e-17;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31462 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 31521

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   31522 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 31562


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 5.3e-17;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25850 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 25909

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25910 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 25950


RESULT 9
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match              64.8%;  Score 65.4;  DB 4;  Length 33529;
  Best Local Similarity    78.8%;  Pred. No. 2.4e-08;
  Matches   78;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
            ||| ||| | |||||||||||| ||||||||||||||||||    |  || ||||||||||
```

```
Db  10351 GCGTGCGGTCTCGCCCGACGGACGGTGCAAGACGTTCTCGCAGGCGGCCGACGGCTACGG 10410

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || |||||||||   ||  |||||||||||| | ||| ||
Db  10411 CCGCGGCGAGGGCGTCGGGGTGGTGGTGCTGATGCGTCT 10449


RESULT 10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             60.8%;  Score 61.4;  DB 4;  Length 4403765;
  Best Local Similarity   77.9%;  Pred. No. 1.8e-07;
  Matches   74;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           | |||  |||||||||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4252406 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4252347

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           ||| ||  ||| | || ||||||||||||| |
Db 4252346 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGGGT 4252312


RESULT 11
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             60.8%;  Score 61.4;  DB 4;  Length 4411529;
  Best Local Similarity   77.9%;  Pred. No. 1.8e-07;
  Matches   74;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           | |||  |||||||||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4260150 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4260091

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           ||| ||  ||| | || ||||||||||||| |
Db 4260090 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGGGT 4260056


RESULT 12
US-09-060-756-318/c
; Sequence 318, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 318
;   LENGTH: 333
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-060-756-318

  Query Match             60.2%;  Score 60.8;  DB 4;  Length 333;
  Best Local Similarity   77.1%;  Pred. No. 4e-07;
  Matches   74;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     182 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 123

Qy      65 GGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           |||||||| |||| ||||||||||||||||||| ||
Db     122 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGGTTG 87


RESULT 13
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1

  Query Match             59.4%;  Score 60;  DB 1;  Length 11219;
  Best Local Similarity   75.0%;  Pred. No. 5.2e-07;
  Matches   75;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           | ||||||| ||||| ||||| ||||||||| | ||||||| ||| || ||||| | ||
Db    7313 GGGCGCGCTCTCGCCGGACGGCCGGTGCAAGCCCTTCTCGGACGAGGCCGACGGATTCGG 7372

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             |  || ||||| |  ||  | || |||||| ||||| ||
Db    7373 TCTCGGGGAGGGTTCGGCGTTCGTCGTGCTCCAGCGGTTG 7412


RESULT 14
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
```

```
;       STRAIN:  NRRL 2338
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  744..6659
;       OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;       OTHER INFORMATION:  MODULE 1"
;       OTHER INFORMATION:  /label= FUNCTION
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  744..11219
;       OTHER INFORMATION:  /function= "gene= "eryA""
;       OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;       OTHER INFORMATION:  6-deoxyerythronolide B""
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  744..1868
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1998..2198
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2250..3626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3831..4811
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5574..6125
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6369..6626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..11219
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..8066
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8262..9305
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9906..10454
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10707..10964
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1

  Query Match             59.4%;  Score 60;  DB 3;  Length 11219;
  Best Local Similarity   75.0%;  Pred. No. 5.2e-07;
  Matches   75;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           | ||||||| ||||| ||||| ||||||||| | ||||||| ||| || ||||| | ||
Db    7313 GGGCGCGCTCTCGCCGGACGGCCGGTGCAAGCCCTTCTCGGACGAGGCCGACGGATTCGG 7372

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             |  || ||||| |  ||  | || |||||| ||||| ||
Db    7373 TCTCGGGGAGGGTTCGGCGTTCGTCGTGCTCCAGCGGTTG 7412


RESULT 15
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Beck, James J.
;    APPLICANT:  Hill, Dwight S.
;    APPLICANT:  Neff, Snezanna
;    APPLICANT:  Ryals, John A.
;    TITLE OF INVENTION:  Genes For The Biosynthesis Of Soraphen
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road, P.O. Box 2005
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/764,233A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/729,214
;      FILING DATE:  09-OCT-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/258,261
;      FILING DATE:  08-JUN-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  1506/CIP6
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (919) 541-8587
;      TELEFAX:  (919) 541-8689
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  49377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Sorangium cellulosum
;    IMMEDIATE SOURCE:
;      CLONE:  p98/1, pJL3, and pVKM15
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  383..760
;      OTHER INFORMATION:  /product= "SorR"
;      OTHER INFORMATION:  /note= "This gene encodes a protein that is highly homolog
;      OTHER INFORMATION:  the reductase domains of type I PKSs such as eryA from
;      OTHER INFORMATION:  Saccharopolyspora erythraea."
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  927..19874
```

```
;       OTHER INFORMATION:  /product= "SorA"
;       OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKSs that
;       OTHER INFORMATION:  are known to be involved in the synthesis of polyketide
;       OTHER INFORMATION:  compounds."
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  942..7115
;       OTHER INFORMATION:  /product= "Module 1 of SorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7203..12884
;       OTHER INFORMATION:  /product= "Module 2 of SorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  13455..19616
;       OTHER INFORMATION:  /product= "Module 3 of SorA"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19871..46318
;       OTHER INFORMATION:  /product= "SorB"
;       OTHER INFORMATION:  /note= "Gene product is highly homologous to type I PKS genes
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19870..24556
;       OTHER INFORMATION:  /product= "Module 1 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  24638..30820
;       OTHER INFORMATION:  /product= "Module 2 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  30881..35446
;       OTHER INFORMATION:  /product= "Module 3 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  35528..40114
;       OTHER INFORMATION:  /product= "Module 4 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  40190..46318
;       OTHER INFORMATION:  /product= "Module 5 of SorB"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  46851..47891
;       OTHER INFORMATION:  /product= "SorM"
;       OTHER INFORMATION:  /note= "The protein encoded by the sorM gene is highly
;       OTHER INFORMATION:  homologous to the methyltransferase from Streptomyces
;       OTHER INFORMATION:  hygroscopicus that is involved in the synthesis of the
;       OTHER INFORMATION:  polyketide rappamicin."
US-08-764-233A-1

  Query Match             59.0%;  Score 59.6;  DB 1;  Length 49377;
  Best Local Similarity   75.5%;  Pred. No. 6.1e-07;
  Matches   74;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       3 GCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGC 62
           |||  || || |||||||| || |||   || |||   ||   |||| ||||||| | ||
Db    1900 GCGGCCTGTCCCCCGACGGCCGCTGCTTCACCTTCGATGCGCGCGCGAACGGCTATGTGC 1959

Qy      63 GGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             | ||||||||| ||| ||||||||||||||| || |||
Db    1960 GCGGCGAGGGGGGCGGCGTGGTGGTGCTCAAACGCCTG 1997


Search completed: November  5, 2002, 15:39:42
Job time : 1663.74 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_25850_25950
Perfect score: 101
Sequence: 1 gcgcgcgctttcgcccgacg.....ggtggtgctcaagcggctcc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56	55.4	15872	10	US-09-861-289-1	Sequence 1, Appli
2	47.8	47.3	13842	10	US-09-861-289-30	Sequence 30, Appl
3	47.8	47.3	36778	10	US-09-861-289-5	Sequence 5, Appli
4	45.6	45.1	11220	10	US-09-861-289-32	Sequence 32, Appl
5	44.6	44.2	4041	10	US-09-861-289-36	Sequence 36, Appl
6	44	43.6	4689	10	US-09-861-289-34	Sequence 34, Appl
7	42.8	42.4	643	10	US-09-924-256A-53	Sequence 53, Appl
8	41.8	41.4	658	10	US-09-924-256A-69	Sequence 69, Appl
9	34.6	34.3	765	10	US-09-924-256A-57	Sequence 57, Appl
10	34.4	34.1	761	10	US-09-924-256A-79	Sequence 79, Appl
11	32.6	32.3	435	10	US-09-960-352-175	Sequence 175, App
12	32.6	32.3	643	10	US-09-924-256A-29	Sequence 29, Appl
13	32.2	31.9	1466	10	US-09-820-893-37	Sequence 37, Appl
14	32.2	31.9	13273	10	US-09-764-869-2349	Sequence 2349, Ap
15	31.4	31.1	722	10	US-09-924-256A-49	Sequence 49, Appl
16	31	30.7	643	10	US-09-924-256A-31	Sequence 31, Appl
17	30.8	30.5	753	10	US-09-924-256A-71	Sequence 71, Appl
18	30.6	30.3	655	10	US-09-924-256A-45	Sequence 45, Appl
c 19	30.6	30.3	3571	10	US-09-823-033-3	Sequence 3, Appli
20	30.2	29.9	369	10	US-09-864-761-31209	Sequence 31209, A
21	30.2	29.9	459	10	US-09-864-761-14676	Sequence 14676, A
22	30.2	29.9	501	10	US-09-833-790-164	Sequence 164, App
23	30.2	29.9	637	10	US-09-924-256A-41	Sequence 41, Appl
24	30.2	29.9	1209	10	US-09-880-107-2104	Sequence 2104, Ap
25	30.2	29.9	3966	10	US-09-764-855-227	Sequence 227, App
26	30	29.7	703	10	US-09-924-256A-51	Sequence 51, Appl
c 27	30	29.7	817	12	US-10-062-254-217	Sequence 217, App
c 28	29.6	29.3	4041	10	US-09-861-289-36	Sequence 36, Appl
c 29	29.6	29.3	36778	10	US-09-861-289-5	Sequence 5, Appli
30	29.2	28.9	649	10	US-09-924-256A-61	Sequence 61, Appl
c 31	29	28.7	1550	12	US-10-044-090-740	Sequence 740, App
32	28.8	28.5	5184	10	US-09-845-583-9	Sequence 9, Appli
c 33	28.6	28.3	347	10	US-09-759-143-725	Sequence 725, App
c 34	28.6	28.3	347	10	US-09-780-669-725	Sequence 725, App
c 35	28.6	28.3	347	10	US-09-822-827-725	Sequence 725, App
36	28.6	28.3	369	10	US-09-864-761-28176	Sequence 28176, A
37	28.6	28.3	454	10	US-09-864-761-11590	Sequence 11590, A
38	28.4	28.1	637	10	US-09-924-256A-21	Sequence 21, Appl
39	28.4	28.1	1152	10	US-09-147-346-1	Sequence 1, Appli
40	28.4	28.1	2760	10	US-09-452-599-20	Sequence 20, Appl
41	28.2	27.9	144	10	US-09-918-063-8	Sequence 8, Appli
c 42	28.2	27.9	144	10	US-09-918-063-10	Sequence 10, Appl
43	28.2	27.9	978	10	US-09-918-063-5	Sequence 5, Appli
c 44	28.2	27.9	978	10	US-09-918-063-7	Sequence 7, Appli
45	28.2	27.9	1236	10	US-09-918-063-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 55.4%; Score 56; DB 10; Length 15872;
Best Local Similarity 74.0%; Pred. No. 6.6e-07;
Matches 71; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

```
Qy       5 GCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 64
           | |||  |  ||||||| || |||||| |||||||||||| ||| ||||||  ||   ||
Db    9146 GGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGG 9205

Qy      65 GGCGAGGGGTGCGCCCGTGGTGGTGCTCAAGCGGCTG 100
             | ||||||   || |||| | || ||  ||||||||
Db    9206 GCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTG 9241
```

RESULT 2
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/851,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             47.3%;  Score 47.8;  DB 10;  Length 13842;
  Best Local Similarity   67.7%;  Pred. No. 8.8e-05;
  Matches   67;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
             ||| |||||  | || ||||| ||  ||||| | |||||||| ||||| ||||| | ||
Db    8454 GCGGGCGCTCGCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGG 8513

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
               |  |||||   || | || || |  |  |||||||
Db    8514 CGCCGCGGAGGGCGTCGGCCTGCTGCTCGTGGAGCGGCT 8552


RESULT 3
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             47.3%;  Score 47.8;  DB 10;  Length 36778;
  Best Local Similarity   67.7%;  Pred. No. 8.3e-05;
  Matches   67;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
             ||| |||||  | || ||||| ||  ||||| | |||||||| ||||| ||||| | ||
Db   10195 GCGGGCGCTCGCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGG 10254

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
               |  |||||   || | || || |  |  |||||||
Db   10255 CGCCGCGGAGGGCGTCGGCCTGCTGCTCGTGGAGCGGCT 10293


RESULT 4
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             45.1%;  Score 45.6;  DB 10;  Length 11220;
  Best Local Similarity   66.0%;  Pred. No. 0.00033;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
             ||||| |||  ||   ||||| ||||  ||| ||||| | ||    ||||||||| ||
Db    5262 GCGCGGGCTCGCGGAGGACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGG 5321

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             | ||  |||||   || | || |  |  || |||| |||
Db    5322 CCCGGCGGAGGGCGTCGGCATGCTCCTCGTCGAGCGCCTG 5361


RESULT 5
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-36

  Query Match             44.2%;  Score 44.6;  DB 10;  Length 4041;
  Best Local Similarity   65.7%;  Pred. No. 0.00063;
  Matches   65;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
             ||  ||   |  |  ||||||| |||  ||||   ||| |  | |  |||||||||  |||
Db     747 GCAGGCCATGGCCGCCGACGGCCGGACCAAGGGCTTCGCCTCGGCGGCGGACGGCCTCGC 806

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             ||||||||||   ||||||  |  | ||| | |||||
Db     807 CTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAACGGCT 845


RESULT 6
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
```

```
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             43.6%;  Score 44;  DB 10;  Length 4689;
  Best Local Similarity   65.0%;  Pred. No. 0.0009;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
             ||||| ||| | | ||||| |||| ||| ||||| | || |||||||||| ||
Db     753 GCGCGGGCTGGCCGCGGACGGCCGGTCGAAGGCGTTCGCCACCTCGGCGGACGGCTTCGG 812

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             | .|  |||||   ||  ||  || || |  |||| |||
Db     813 CCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTG 852


RESULT 7
US-09-924-256A-53
; Sequence 53, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-09-924-256A-53

  Query Match             42.4%;  Score 42.8;  DB 10;  Length 643;
  Best Local Similarity   68.6%;  Pred. No. 0.002;
  Matches   59;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      15 CCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGGGT 74
           | || || || ||||| || ||    || || || || || |||| |||||||||| || |
Db     384 CTGATGGTCGTTGCAAAACCTTTGATGCTGATGCTGATGGTTACGTGCGGGGCGAAGGTT 443

Qy      75 GCGCCGTGGTGGTGCTCAAGCGGCTG 100
           | |  || ||  | ||||||||  ||
Db     444 GTGGTGTTGTAATTCTCAAGCGTTTG 469


RESULT 8
US-09-924-256A-69
; Sequence 69, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
;   LENGTH: 658
;   TYPE: DNA
;   ORGANISM: Usnea florida
US-09-924-256A-69

  Query Match             41.4%;  Score 41.8;  DB 10;  Length 658;
  Best Local Similarity   65.6%;  Pred. No. 0.0037;
  Matches   61;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       7 GCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGG 66
           |||||| ||    || || | |   | ||     || |||||||| |||||||| || ||
Db     376 GCTTTCTCCTACGGGCCGCTCCCGCATGTGGGATGCGGACGCGGATGGCTACGCTCGTGG 435

Qy      67 CGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
             |||||    || |  || ||||| || |||||
Db     436 GGAGGGCGTAGCGGCTGTAGTGCTGAAACGGCT 468


RESULT 9
US-09-924-256A-57
; Sequence 57, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
;   LENGTH: 765
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-09-924-256A-57

  Query Match             34.3%;  Score 34.6;  DB 10;  Length 765;
  Best Local Similarity   59.8%;  Pred. No. 0.27;
  Matches   58;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       3 GCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGC 62
           | ||||| || || || ||   |||| | || |||   |     ||  | || ||  | |
Db     494 GAGCGCTCTCTCCTGATGGCGTGTGCTACACTTTCGATGAACGGGCCAATGGATATTCCC 553

Qy      63 GGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           |||| || ||   || |     | || ||||| |||||
Db     554 GGGGAGAAGGTGTCGGCACCATCGTTCTCAAACGGCT 590


RESULT 10
US-09-924-256A-79
; Sequence 79, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
```

```
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
;   LENGTH: 761
;   TYPE: DNA
;   ORGANISM: Usnea florida
US-09-924-256A-79

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 761;
  Best Local Similarity   60.9%;  Pred. No. 0.3;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       8 CTTTCGCCCGACGGGCGGTGCAAGACGTTC TCGGCCGACGCGGACGGCTACGCGCGGGGC 67
           || || |||||| | || |     ||  ||| |  | || || ||||| ||
Db     495 CTGTCTCCCGACAGCCGTTCTTGGAGTTTCGATCAACGTGCTAATGGTTATGCGCGTGGG 554

Qy      68 GAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCT 99
           || ||   ||||||| |||| |||| || ||
Db     555 GAAGGAACCGCCGTGATGGTCGTCAAACGCCT 586


RESULT 11
US-09-960-352-175
; Sequence 175, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960 352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 175
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (386)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: 01-LIB34-045-Q1-E1-A1
US-09-960-352-175

  Query Match             32.3%;  Score 32.6;  DB 10;  Length 435;
  Best Local Similarity   63.3%;  Pred. No. 0.91;
  Matches   50;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      23 CGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGGGTGCGCCGTG 82
           | ||||  |  |||||||| ||||| ||    |  ||||||| ||||| | | |  ||
Db     186 CTGTGCCGGCGGTTCTCGGGCGACGTGGTGCTCACGGCGCGGGACGAGGCGCGGGGTCTG 245

Qy      83 GTGGTGCTCAAGCGGCTGC 101
           | ||   |  |||  ||||
Db     246 GCGGCCGTGCAGCATCTGC 264


RESULT 12
US-09-924-256A-29
; Sequence 29, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Clone
;   OTHER INFORMATION: Xa.cum.6H
US-09-924-256A-29

  Query Match             32.3%;  Score 32.6;  DB 10;  Length 643;
  Best Local Similarity   60.9%;  Pred. No. 0.89;
  Matches   53;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      14 CCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGGG 73
           ||   |||||  |||||| |||     ||    || || ||||||   || || || ||
Db     383 CCTAGCGGGCAATGCAAGCCGTTTGATGCTTCTGCTGATGGCTACTGCCGTGGAGAAGGA 442

Qy      74 TGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             |  || || ||||| ||||   ||
Db     443 GTTGGTGTCGTTGTGCTTAAGCCTTTG 469


RESULT 13
US-09-820-893-37
; Sequence 37, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 31 Human Secreted Proteins
;  FILE REFERENCE: PZ033P1
;  CURRENT APPLICATION NUMBER: US/09/820,893
;  CURRENT FILING DATE:  2001-03-30
;  PRIOR APPLICATION NUMBER: 09/531,119
;  PRIOR FILING DATE: 2000-03-20
;  PRIOR APPLICATION NUMBER: 60/102,895
;  PRIOR FILING DATE: 1998-10-02
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 37
;   LENGTH: 1466
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-820-893-37

  Query Match             31.9%;  Score 32.2;  DB 10;  Length 1466;
  Best Local Similarity   61.2%;  Pred. No. 1.1;
  Matches   52;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      16 CGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGGGTG 75
           || ||      | ||||| || |   |  |||    || |||||  | | ||||| |||
Db     443 CGCCGCCTCCTTCAAGAAGTACCACGTGGACCACCACCGCTACCTGGGCGGCGACGGGCT 502

Qy      76 CGCCGTGGTGGTGCTCAAGCGGCTG 100
             | |||||  |||| || ||| |||
Db     503 GGACGTGGACGTGCCCACGCGTCTG 527


RESULT 14
US-09-764-869-2349
```

```
; Sequence 2349, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007
;  CURRENT APPLICATION NUMBER: US/09/764,869
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 2442
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2349
;   LENGTH: 13273
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-2349

  Query Match             31.9%;  Score 32.2;  DB 10;  Length 13273;
  Best Local Similarity   57.4%;  Pred. No. 0.95;
  Matches   58;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           || |  |||  ||  | | | || ||   |   | | |  |   |||||| |||||||||
Db    1884 GCCCCGGCTCCCGTTCCAGGTGCTGTTGCACGTGCTGTTTGAGCACGCGGTCGGCTACGC 1943

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           || |   | |  |   |  |||| || | |||  | |||||
Db    1944 GCTGCTGGCGCTGAAGGAAGTGGAGGAGATCAGTCTGCTGC 1984


RESULT 15
US-09-924-256A-49
; Sequence 49, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 722
;   TYPE: DNA
;   ORGANISM: Parmelia sulcata
US-09-924-256A-49

  Query Match             31.1%;  Score 31.4;  DB 10;  Length 722;
  Best Local Similarity   57.5%;  Pred. No. 1.8;
  Matches   50;  Conservative    3;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       7 GCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGG 66
           | | || ||    ||:||:| |   | ||     ||  ||:|||| ||||| || || ||
Db     443 GTTGTCCCCAACTGGSCGSTCCCGAATGTGGGACGCTAACSCGGATGGCTATGCACGAGG 502

Qy      67 CGAGGGGTGCGCCGTGGTGGTGCTCAA 93
            |||||    ||    || || |||||
Db     503 AGAGGGTATTGCATCTGTCGTACTCAA 529


Search completed: November  5, 2002, 23:01:17
Job time : 97.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_25850_25950
Perfect score: 101
Sequence: 1 gcgcgcgctttcgcccgacg.....ggtggtgctcaagcggctgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	41.8	41.4	516	9	AI532314	AI532314 SD03777.5
	2	40.8	40.4	597	13	BM488815	BM488815 pgm2n.pk0
c	3	36.2	35.8	738	17	BH179086	BH179086 013_D_19-
c	4	36.2	35.8	738	17	CNS07L94	AL616042 T3 end of
	5	35.4	35.0	898	12	BF182804	BF182804 601809319
	6	35.4	35.0	1254	14	BM911669	BM911669 AGENCOURT
c	7	35	34.7	1580	17	AG083426	AG083426 Pan trogl
	8	34.4	34.1	1188	14	BM908892	BM908892 AGENCOURT
	9	34.2	33.9	463	17	BH611949	BH611949 SALK_0319
	10	34	33.7	470	12	BF429029	BF429029 WHE1704_C
c	11	34	33.7	559	9	AL564234	AL564234 AL564234
c	12	34	33.7	775	17	AZ185516	AZ185516 SP_1005_A
c	13	34	33.7	1096	14	BM910915	BM910915 AGENCOURT
c	14	33.8	33.5	155	14	BQ468858	BQ468858 HM02J05r
	15	33.8	33.5	660	12	BE903754	BE903754 601675594
	16	33.6	33.3	932	17	CNS0072Q	AL066742 Drosophil
	17	33.4	33.1	442	14	BM754672	BM754672 K-EST0032
	18	33.4	33.1	531	14	BQ245189	BQ245189 TaE15029D
c	19	33.4	33.1	1582	17	AG054236	AG054236 Pan trogl
	20	33.2	32.9	721	12	BG698427	BG698427 602658446
	21	33.2	32.9	934	14	BQ674384	BQ674384 AGENCOURT
	22	33.2	32.9	1219	14	BM803425	BM803425 AGENCOURT
c	23	33	32.7	504	14	BQ514230	BQ514230 EST621645
c	24	33	32.7	757	13	BM113440	BM113440 EST560976
c	25	33	32.7	937	17	AG080291	AG080291 Pan trogl
	26	33	32.7	945	14	BQ677554	BQ677554 AGENCOURT
	27	33	32.7	1152	14	BM925325	BM925325 AGENCOURT
	28	32.8	32.5	478	14	BM835032	BM835032 K-EST0110
	29	32.8	32.5	605	14	BM789564	BM789564 K-EST0069
	30	32.8	32.5	656	13	BI951962	BI951962 HVSMEm000
	31	32.8	32.5	873	17	AG043471	AG043471 Pan trogl
c	32	32.8	32.5	897	17	AG046101	AG046101 Pan trogl
	33	32.8	32.5	992	12	BG288191	BG288191 602383708
c	34	32.8	32.5	1819	14	BQ956172	BQ956172 AGENCOURT
	35	32.6	32.3	611	10	AV611810	AV611810 AV611810
	36	32.6	32.3	1134	14	BQ676007	BQ676007 AGENCOURT
c	37	32.6	32.3	1682	12	BF782466	BF782466 602106175
	38	32.6	32.3	1798	17	AG171124	AG171124 Pan trogl
c	39	32.4	32.1	390	9	AU070752	AU070752 AU070752
c	40	32.4	32.1	413	9	AU163463	AU163463 AU163463
c	41	32.4	32.1	970	12	BG855445	BG855445 1024042G0
c	42	32.4	32.1	1057	17	CNS046BO	AL276477 Tetraodon
	43	32.2	31.9	429	10	BE355032	BE355032 DG1_10_A0
	44	32.2	31.9	509	12	BG739820	BG739820 602630546
	45	32.2	31.9	546	14	BQ129385	BQ129385 ij35d01.y

ALIGNMENTS

RESULT 1
AI532314
LOCUS AI532314 516 bp mRNA linear EST 19-APR-2001
DEFINITION SD03777.5prime SD Drosophila melanogaster Schneider L2 cell culture pOT2 Drosophila melanogaster cDNA clone SD03777 5prime, mRNA sequence.
ACCESSION AI532314
VERSION AI532314.1 GI:4446449
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 516)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AC005554
Plate: 37 row: G column: 5
High quality sequence stop: 514.
FEATURES Location/Qualifiers

```
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD03777"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      101 a    163 c    158 g     94 t
ORIGIN

  Query Match             41.4%;  Score 41.8;  DB 9;  Length 516;
  Best Local Similarity   68.2%;  Pred. No. 2;
  Matches   58;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      14 CCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGGG 73
           | ||||||    ||||||||| |||   || |  ||  ||||||| || ||||  ||
Db     201 CACGACGGCTCCTGCAAGACCTTCGATGCCGGCGGCCAACGGCTATGCCCGGGCGGACACC 260

Qy      74 TGCGCCGTGGTGGTGCTCAAGCGGC 98
           ||||||||| ||| ||||  || |||
Db     261 TGCGCCGTTGTGCTGCTTCAGAGGC 285


RESULT 2
BM488815
LOCUS       BM488815                 597 bp    mRNA    linear   EST 07-FEB-2002
DEFINITION  pgm2n.pk008.p20 Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA
            clone pgm2n.pk008.p20 5' similar to ref|NP_446119.1 (NM_053667)
            growth supressor 1; leprecan [Rattus norvegicus]
            gb|AAD51875.1|AF087433_1 (AF087433) leprecan [Rattus norvegicus],
            mRNA sequence.
ACCESSION   BM488815
VERSION     BM488815.1  GI:18609746
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Cogburn,L.A. and Monsonego-Ornan,E.
  TITLE     ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome
            Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Gallus gallus"
                     /strain="Commercial broiler and  Ottawa Res. Centre
                     Strains 90 & 21"
                     /db_xref="taxon:9031"
                     /clone="pgm2n.pk008.p20"
                     /clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
                     and Epiphyseal Growth Plate cDNA library (pgm2n)"
                     /sex="Male and Female"
                     /tissue_type="Breast muscle, leg muscle and epiphyseal
                     growth plate"
                     /dev_stage="Breast,leg:Embryo(d19);post-hatch(1d,1,3,5,7,9
                     ,11 weeks);growth plate(1d,7d,14d post-hatch)"
                     /lab_host="E. coli ElDH10B"
                     /note="Vector: pCMVSPORT6; Library made from equivalent
                     pools of total RNA isolated from each tissue (embryonic
                     muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
                     plate 33.3% of the final RNA pool). Single pass sequencing
                     from 5'-end"
BASE COUNT      78 a    200 c    199 g     90 t     30 others
ORIGIN

  Query Match             40.4%;  Score 40.8;  DB 13;  Length 597;
  Best Local Similarity   63.0%;  Pred. No. 3.4;
  Matches   63;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           || | | ||  || ||||    | |  |  |   ||| | |||| ||| || |  |||||
Db      68 GCCCCCTCTGCCGGCCGAGCCCCCGGACGCGCTCTTCGCTGCCGGCGCCGAGGCGTACGC 127

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
           |||||| ||  ||  |||||||||| |||  | |  || |
Db     128 GCGGGGGGACTGGCCCGCCGTGGTGCTGCAGATGGAGCGG 167


RESULT 3
BH179086/c
LOCUS       BH179086                 738 bp    DNA     linear   GSS 19-OCT-2001
DEFINITION  013_D_19-rev SmBAC1 Schistosoma mansoni genomic clone 013D19 5',
            DNA sequence.
ACCESSION   BH179086
VERSION     BH179086.1  GI:16279565
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Le Paslier,M.-C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams
            ,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
COMMENT     Other_GSSs: 013_D_19-21
            Contact: Pierce RJ
            INSERM U 167
            Institut Pasteur de Lille
            1 rue du Professeur A. Calmette, 59019-Lille, France
            Tel: (33) (0)3 20877783
            Fax: (33) (0)3 20877888
            Email: Raymond.Pierce@pasteur-lille.fr
            CNS sequencing ID=DG0AA013CB10BP1
            Plate: 013  row: D  column: 19
            Seq primer: M13 reverse primer
            Class: BAC ends
            High quality sequence stop: 738.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Schistosoma mansoni"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="013D19"
                     /clone_lib="SmBAC1"
                     /sex="mixed"
                     /dev_stage="cercariae"
                     /lab_host="Biomphalaria glabrata"
                     /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                     Hind III digested and size-selected S. mansoni  cercarial
                     DNA was ligated into Hind III digested pBeloBAC 11 vector
                     and used to transform E. coli DH10B. The complete library
                     contains 23808 clones from 4 independent
                     sizing-ligation-transformations.  Average insert size
                     ranges from 70-127 kb and genome coverage is 7.9-fold."
BASE COUNT      239 a    147 c    124 g    168 t     60 others
ORIGIN

  Query Match             35.8%;  Score 36.2;  DB 17;  Length 738;
  Best Local Similarity   32.7%;  Pred. No. 37;
  Matches   33;  Conservative   38;  Mismatches   30;  Indels    0;  Gaps    0;
```

```
Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ::|:|:|: | ||: :| :::::| :   | ||  :  ||::: :: |:  ||:   :
Db    118 SSGSGSGSSTCCGSGSGSSSSSSSGGSGGGGTCGGGSSGGGSSSSSSGGSSGGGSSGGSG 59

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          | |:||:| |||| |::::| :|::|:  |   ||::::|:
Db     58 GGGSGGSGSGGGGGGSSSSGSKGKKGKTGTTTCGCSBBBGB 18


RESULT 4
CNS07L94/c
LOCUS       CNS07L94                738 bp    DNA     linear   GSS 08-OCT-2001
DEFINITION  T3 end of clone 013CB10 of library SmBAC1 from strain Puerto-Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL616042
VERSION     AL616042.1  GI:16029266
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni.
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 738)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Partially Hind III digested and size-selected S. mansoni cercarial
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing-ligation-transformations. Average
            insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Schistosoma mansoni"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="013CB10"
                     /clone_lib="SmBAC1"
                     /note="end : T3"
BASE COUNT      239 a    147 c    124 g    168 t     60 others
ORIGIN

  Query Match             35.8%;  Score 36.2;  DB 17;  Length 738;
  Best Local Similarity   32.7%;  Pred. No. 37;
  Matches   33;  Conservative   38;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ::|:|:|: | ||: :| :::::| :   | ||  :  ||::: :: |:  ||:   :
Db    118 SSGSGSGSSTCCGSGSGSSSSSSSGGSGGGGTCGGGSSGGGSSSSSSGGSSGGGSSGGSG 59

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          | |:||:| |||| |::::| :|::|:  |   ||::::|:
Db     58 GGGSGGSGSGGGGGGSSSSGSKGKKGKTGTTTCGCSBBBGB 18


RESULT 5
BF182804
LOCUS       BF182804                898 bp    mRNA    linear   EST 31-OCT-2000
DEFINITION  601809319F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040045 5',
            mRNA sequence.
ACCESSION   BF182804
VERSION     BF182804.1  GI:11060947
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM857  row: e  column: 06
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4040045"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      203 a    219 c    293 g    183 t
ORIGIN

  Query Match             35.0%;  Score 35.4;  DB 12;  Length 898;
  Best Local Similarity   59.4%;  Pred. No. 55;
  Matches   60;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy      1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
          ||| | |||  ||||     |  | |||| |  | | |  |   ||||| |||||||||
Db      1 GCGGGAGCTGGCGCCATGGTGCTGTTGCACGGTGCTGTTTGAGCACGCGGTCGGCTACGC 60

Qy     61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
          || |   | |  |   |  |||| || | |||  | |||||
Db     61 GCTGCTGGCGCTGAAGGAAGTGGAGGAGATCAGTCTGCTGC 101


RESULT 6
BM911669
LOCUS       BM911669                1254 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6615158 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466497
            5', mRNA sequence.
ACCESSION   BM911669
VERSION     BM911669.1  GI:19362048
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1254)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
```

```
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1966  row: p  column: 18
            High quality sequence stop: 396.
FEATURES              Location/Qualifiers
     source           1. .1254
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5466497"
                      /clone_lib="NIH_MGC_41"
                      /tissue_type="amelanotic melanoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                      into EcoRI/XhoI sites using the following 5' adaptor:
                      GGCACGAG(G). Library constructed by Ling Hong in the
                      laboratory of Gerald M. Rubin (University of California,
                      Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                      Superscript II RT (Life Technologies). Note: this is a
                      NIH_MGC Library."
BASE COUNT      156 a    482 c    480 g    130 t      6 others
ORIGIN

  Query Match             35.0%;  Score 35.4;  DB 14;  Length 1254;
  Best Local Similarity   59.4%;  Pred. No. 54;
  Matches   60;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||| ||||   || ||| | ||| | |   |  | || | | ||||| |||    ||
Db     466 GCGGGCGCCGCCGTCCGCCTGGCCGCGGCGCTGGCCGCGTCGGGCGCGGCCGGGCCGGC 525

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           | | ||   |||| | || |  | || || |  ||||| ||
Db     526 GGGCGGGCGGGGGGGGGCGGGCGGGGGGCGCCGGCGGCGGC 566


RESULT 7
AG083426/c
LOCUS       AG083426               1580 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-080N21.R, genomic survey sequence.
ACCESSION   AG083426
VERSION     AG083426.1  GI:16635228
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-080N21.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata  Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates, Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1580)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES              Location/Qualifiers
     source           1. .1580
                      /organism="Pan troglodytes"
                      /db_xref="taxon:9598"
                      /clone="PTB-080N21.R"
                      /sex="male"
                      /cell_type="lymphoblast"
                      /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      277 a    584 c    609 g     55 t      55 others
ORIGIN

  Query Match             34.7%;  Score 35;  DB 17;  Length 1580;
  Best Local Similarity   63.3%;  Pred. No. 65;
  Matches   50;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           || ||||    | ||| |||| || ||  | || |  ||||   ||||| |     ||||
Db     299 CGGGCGCGCCGGNCCGNCGGGGGGCGCGGGGCGCTGCCGGCGCCCGCGGCCNNGGGCGCG 240

Qy      62 CGGGGCGAGGGGTGCGCCG 80
           | |  || |||| ||||||
Db     239 CCGCCCGCGGGGCGCGCCG 221


RESULT 8
BM908892
LOCUS       BM908892               1188 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6611077 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431496
            5', mRNA sequence.
ACCESSION   BM908892
VERSION     BM908892.1  GI:19359271
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1188)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1903  row: n  column: 09
            High quality sequence stop: 438.
FEATURES              Location/Qualifiers
     source           1. .1188
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:5431496"
                      /clone_lib="NIH_MGC_98"
                      /tissue_type="astrocytoma grade IV, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                      EcoRI; cDNA made by oligo-dT priming.  Directionally
                      cloned into EcoRI/XhoI sites using the following 5'
                      adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                      in the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies).
                      Note: this is a NIH_MGC Library."
BASE COUNT       57 a    503 c    416 g    202 t      10 others
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 14;  Length 1188;
  Best Local Similarity   65.8%;  Pred. No. 91;
  Matches   50;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;
```

```
Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           ||||||||   || ||| || |||| ||    |  |||||||| ||     |||  | |
Db     785 GCGCGCGCCGCCGGCCGTCGCGCGGGGCGGCGCCCTCTCGGCCGGCGGTCCGGGCGCCCC 844

Qy      61 GCGGGGCGAGGGGTGC 76
           |||||  | | |||||
Db     845 GCGGGCGGCGCGGTGC 860


RESULT 9
BH611949
LOCUS       BH611949                463 bp    DNA     linear   GSS 04-JAN-2002
DEFINITION  SALK_031928 Arabidopsis thaliana TDNA insertion lines Arabidopsis
            thaliana genomic clone SALK_031928, DNA sequence.
ACCESSION   BH611949
VERSION     BH611949.1  GI:18059245
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 463)
  AUTHORS   Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
            ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
            , Zimmerman,J. and Ecker,J.R.
  TITLE     A Sequence-Indexed Library of Insertion Mutations in the
            Arabidopsis Genome
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Joseph R. Ecker
            Salk Institute Genomic Analysis Laboratory (SIGnAL)
            The Salk Institute for Biological Studies
            10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
            Tel: 858 453 4100 x1752
            Fax: 858 558 6379
            Email: ecker@salk.edu
            This is single pass sequence recovered from the left border of
            TDNA.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .463
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia 0"
                     /db_xref="taxon:3702"
                     /clone="SALK_031928"
                     /clone_lib="Arabidopsis thaliana TDNA insertion lines"
                     /note="PCR was performed on Arabidopsis thaliana lines
                     each of which contains  one or more TDNA insertion
                     elements. The resultant fragment for each  line was
                     directly sequenced to determine the genomic sequence at
                     the  site of insertion. Details of the protocols used can
                     be found at  http://signal.salk.edu/tdna_protocols.html"
BASE COUNT       63 a    111 c    190 g     71 t     28 others
ORIGIN

  Query Match            33.9%;  Score 34.2;  DB 17;  Length 463;
  Best Local Similarity  59.4%;  Pred. No. 1.1e+02;
  Matches   57;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       3 GCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGC 62
           | ||||   |||||| |||||||  |   | ||     ||| | ||| |   |   ||  |
Db     351 GGGCGCGCCCGCCCGCCGGGCGGGGGGGGGCGGGGGGGGCGGGCGCCGGGTGGGCCGGCC 410

Qy      63 GGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGC 98
           ||||   |||| ||| || || || |     |||||
Db     411 GGGGGCGGGGGGGCGGCGCGGGGGCGNGGCGGCGGC 446


RESULT 10
BF429029
LOCUS       BF429029                470 bp    mRNA    linear   EST 29-NOV-2000
DEFINITION  WHE1704_C03_F06ZS Wheat heat stressed spike cDNA library Triticum
            aestivum cDNA clone WHE1704_C03_F06, mRNA sequence.
ACCESSION   BF429029
VERSION     BF429029.1  GI:11440948
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
            Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Heat stressed spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1704_C03_F06"
                     /clone_lib="Wheat heat stressed spike cDNA library"
                     /tissue_type="Whole spike"
                     /dev_stage="Spikes at 5, 10, 15 and 20 days after
                     anthesis"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
                     days after anthesis were heat stressed under  two
                     conditions at Texas Tech University (D. Zhang in HT Nguyrn
                     lab):  (1) at 38 C for 4 hours and (2) 5 days of cyclic
                     treatment of 38 C  for 4 hours.  Total RNA and poly(A) RNA
                     were prepared, a cDNA library  was made, and the cDNA
                     clones were in vivo excised to give pBluescript  phagemids
                     in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
                     the  University of California, Riverside.  Plasmid DNA
                     preparations and  DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT       46 a    220 c    138 g     65 t      1 others
ORIGIN

  Query Match            33.7%;  Score 34;  DB 12;  Length 470;
  Best Local Similarity  58.6%;  Pred. No. 1.2e+02;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
           | |||   | ||     |||| || ||| |  | || ||  ||||||   || || ||||
Db     350 CACGCCGCTGCGGTGCGCGGGGGGAGCANGTGGCTCGCGAGCGACGCCTCCGCCTCCGCG 409

Qy      62 CGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTG 100
             | | |||   | ||||| | |||||     ||  |||
Db     410 GCGAGGGAGCCCTCCGCCGAGCTGGTGGAGGTGCCCCTG 448


RESULT 11
AL564234/c
LOCUS       AL564234                559 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL564234 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM003YF10 3
            prime, mRNA sequence.
ACCESSION   AL564234
VERSION     AL564234.1  GI:12914435
KEYWORDS    EST.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 559)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .559
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DM003YF10"
                     /clone_lib="LTI_NFL001_NBC4"
                     /sex="male"
                     /tissue_type="neuroblastoma cells"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector.  Library was normalized. Library was constructed
                     by Life Technologies.  Contact : Feng Liang Life
                     Technologies, a division of Invitrogen 9800 Medical Center
                     Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
                     8371 Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       69 a    327 c     97 g     44 t     22 others
ORIGIN

  Query Match             33.7%;  Score 34;  DB 9;  Length 559;
  Best Local Similarity   57.4%;  Pred. No. 1.2e+02;
  Matches   58;  Conservative    1;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
           |  |||||   ||| :| |||| || ||    ||   | ||  |  | ||  ||    |
Db     440 GGNCGCGCGCGCGCGSGGCGGGGGGCGCGGCGCGGGTGGGGGGGGGGGGTGGGGGGGG 381

Qy      61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
           |||||| | |||| | |  | || || || |  | ||| ||
Db     380 GCGGGGGGGGGGGGGGGGGCGGGGGCGCGGGGGGCGGC 340


RESULT 12
AZ185516/c
LOCUS       AZ185516                 775 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1005_A1_H12_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
            genomic clone Plate=1005 Col=23 Row=O, DNA sequence.
ACCESSION   AZ185516
VERSION     AZ185516.1  GI:8357801
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 775)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1005  row: O  column: 23
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 775.
FEATURES             Location/Qualifiers
     source          1. .775
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1005 Col=23 Row=O"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT       46 a    451 c    243 g     27 t      8 others
ORIGIN

  Query Match             33.7%;  Score 34;  DB 17;  Length 775;
  Best Local Similarity   66.2%;  Pred. No. 1.2e+02;
  Matches   49;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      13 GCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCGAGGG 72
           | | | ||||||| ||  |  |   |||| || ||||| ||||   | | |||||| |||
Db     303 GGCGGCCGGGCGGCGCGGGCGGGGGTCGGGCGCCGCGGCCGGCGGGGGGGGGGGCGCGGG 244

Qy      73 GTGCGCCGTGGTGG 86
           | | || | || ||
Db     243 GCGGGCGGGGGGGG 230


RESULT 13
BM910915/c
LOCUS       BM910915                 1096 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6615874 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454494
            5', mRNA sequence.
ACCESSION   BM910915
VERSION     BM910915.1  GI:19361294
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1096)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1949  row: l  column: 15
            High quality sequence stop: 590.
FEATURES             Location/Qualifiers
     source          1. .1096
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454494"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
```

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 221 a 380 c 304 g 190 t 1 others
ORIGIN

Query Match 33.7%; Score 34; DB 14; Length 1096;
Best Local Similarity 66.2%; Pred. No. 1.1e+02;
Matches 49; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 2 CGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 61
Db 628 CGCTTGGTCTCGTCTGCTGGGAGGTGGAAGAAGTCCCCAGCCAGTTCGAACTGCGAGGCC 569

Qy 62 CGGGGCGAGGGGTG 75
Db 568 AGGGGCGAGGGCTG 555

RESULT 14
BQ468858/c
LOCUS BQ468858 155 bp mRNA linear EST 30-MAY-2002
DEFINITION HM02J05r HM Hordeum vulgare cDNA clone HM02J05 5-PRIME, mRNA sequence.
ACCESSION BQ468858
VERSION BQ468858.1 GI:21276640
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 155)
AUTHORS Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
TITLE EST sequencing and analysis in barley (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 155 Std Error: 0.00
Plate: 2 row: J column: 5
Seq primer: M13rev.
FEATURES Location/Qualifiers
source 1. .155
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HM02J05"
/clone_lib="HM"
/tissue_type="male inflorescences"
/dev_stage="male inflorescences (approx. 2 mm in size), green anther stage"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRIadapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI,PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinats is not 100% reliable."
BASE COUNT 14 a 62 c 65 g 14 t
ORIGIN

Query Match 33.5%; Score 33.8; DB 14; Length 155;
Best Local Similarity 62.4%; Pred. No. 1.5e+02;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 14 CCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGGCCAGGGG 73
Db 101 CCCGAGCGCGTGTGGCGGGTGTGCACCGGGGACGCCGACGGGGACCCGCCCGGCGACGTC 42

Qy 74 TGCGCCGTGGTGGTGCTCAAGCGGC 98
Db 41 TGCGCGCTGCTGGCGCCGTAGCGCC 17

RESULT 15
BE903754
LOCUS BE903754 660 bp mRNA linear EST 29-SEP-2000
DEFINITION 601675594F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3958415 5', mRNA sequence.
ACCESSION BE903754
VERSION BE903754.1 GI:10395293
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 660)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM837 row: k column: 24
High quality sequence start: 3
High quality sequence stop: 660.
FEATURES Location/Qualifiers
source 1. .660
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3958415"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 145 a 163 c 200 g 152 t
ORIGIN

Query Match 33.5%; Score 33.8; DB 12; Length 660;
Best Local Similarity 58.4%; Pred. No. 1.3e+02;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Qy 1 GCGCGCGCTTTCGCCCGACGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 60
Db 9 GGGGGAGCTGGCGCCATGGTGCTGTTGCACGGTGCTGTTTGAGCACGCGGTCGGCTACGC 68

Qy 61 GCGGGGCGAGGGGTGCGCCGTGGTGGTGCTCAAGCGGCTGC 101
Db 69 GCTGCTGGCGCTGAAGGAAGTGGAGGAGATCAGTCTGCTGC 109

Search completed: November 5, 2002, 16:18:09
Job time : 1117.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_27200_27300
Perfect score: 101
Sequence: 1 cgcagcgtccgaggccgagg.....cgccgtcgtgatcgctggcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	58733	1	AF217189	AF217189 Sorangium
2	101	100.0	71989	6	AR172664	AR172664 Sequence
3	97.8	96.8	68750	1	AF210843	AF210843 Sorangium
4	97.8	96.8	68750	6	AR193029	AR193029 Sequence
5	97.8	96.8	68750	6	AR199551	AR199551 Sequence
6	97.8	96.8	68750	6	AR199559	AR199559 Sequence
7	97.8	96.8	68750	6	AR199567	AR199567 Sequence
8	97.8	96.8	68750	6	AR201097	AR201097 Sequence
9	97.8	96.8	68750	6	AR208671	AR208671 Sequence
10	62.6	62.0	16124	1	AX024384	AX024384 Sequence
11	62.6	62.0	16124	6	AX024277	AX024277 Sequence
c 12	52.2	51.7	26195	1	SC1G7	AL591083 Streptomy
13	50.6	50.1	27541	6	AX211706	AX211706 Sequence
14	50.6	50.1	123580	1	AF263912	AF263912 Streptomy
15	50.6	50.1	125401	6	AX211739	AX211739 Sequence
c 16	49.8	49.3	84985	1	SNA278573	AJ278573 Streptomy
17	47.8	47.3	66808	1	SAU421825	AJ421825 Stigmatel
18	47.4	46.9	113193	1	AF357202	AF357202 Streptomy
19	45.8	45.3	3900	1	AF262754	AF262754 Amycolato
20	45.6	45.1	53784	1	AMM223012	AJ223012 Amycolato
21	45.6	45.1	53789	6	A69720	A69720 Sequence 3
22	45.6	45.1	90445	1	AF040570	AF040570 Amycolato
23	43.4	43.0	5544	1	SNA132221	AJ132221 Streptomy
24	43.4	43.0	5544	6	AX067994	AX067994 Sequence
25	43.4	43.0	84985	1	SNA278573	AJ278573 Streptomy
26	42.6	42.2	78210	1	AB070949	AB070949 Streptomy
27	41.2	40.8	11219	1	SERERYAA	M63676 Saccharopol
28	41.2	40.8	11219	6	AR049367	AR049367 Sequence
29	41.2	40.8	11219	6	AR095528	AR095528 Sequence
30	41	40.6	1404	1	AF015823	AF015823 Streptomy
31	41	40.6	65140	6	AX211705	AX211705 Sequence
32	40.8	40.4	104326	1	AB070940	AB070940 Streptomy
c 33	40.8	40.4	189050	1	AL646066	AL646066 Ralstonia
34	40.6	40.2	42603	1	AF188287	AF188287 Stigmatel
c 35	40.2	39.8	907	8	HVGRP3	Z48624 H.vulgare m
36	40.2	39.8	14402	1	AF098795	AF098795 Pseudomon
c 37	39.8	39.4	39314	1	SGR300302	AJ300302 Streptomy
38	39.4	39.0	43280	1	SFU78289	U78289 Streptomyce
c 39	39.2	38.8	13461	14	AF084543	AF084543 Tupaia he
40	39.2	38.8	195859	14	AF281817	AF281817 Tupaia he
41	38.8	38.4	182569	2	AC130732	AC130732 Oryza sat
c 42	38.6	38.2	1028	6	AR137228	AR137228 Sequence
c 43	38.6	38.2	1028	6	AR156940	AR156940 Sequence
44	38.6	38.2	20235	1	SERERYAB	M63677 S.erythraea
45	38.6	38.2	20235	6	AR049368	AR049368 Sequence

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEIGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIVADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLEPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPLERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRPPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260. .10492
                     /gene="epoB"
     CDS             6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
     gene            10639. .16137
                     /gene="epoC"
     CDS             10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSNAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match           100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 5.6e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27200 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 27259

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             |||||||||||||||||||||||||||||||||||||||||
Db   27260 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 27300


RESULT 2
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 5.3e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27200 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 27259

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             |||||||||||||||||||||||||||||||||||||||||
Db   27260 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 27300


RESULT 3
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
```

/db_xref="GI:6724[illegible]39"
/translation="MGA[illegible]ISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
AGQMPGAVWLVARGDDV[illegible]VDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
EGKLDLDSPVDRWLPEL[illegible]NRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
SSPIQRAIDELGLVNAQF[illegible]VPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
VGRAADQGFDAFVRERI[illegible]APLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
GAESAYASPPAFPSGAAG[illegible]LVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
TPAQKAASSFFPGFFETH[illegible]GWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
IGIVMTQSAGFLFSGALE[illegible]RFWRSVYVATESA"
CDS 3415. .5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26[illegible]06.1"
/db_xref="GI:672424[illegible]0"
/translation="MHGL[illegible]ERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
GVVLGPSVVGALAPGFHR[illegible]LFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
RPGALSALGAIAPPLAAG[illegible]AFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
RRSYAQVTLAAGVVSEVA[illegible]WVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
LTHLAMRWVADATRVSKG[illegible]VSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
NRPLLDGVQTLVAGLFAP[illegible]FFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
LGARLGGLRGSEAALVAV[illegible]LNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
SPALLIWLEKRAPPTQEE[illegible]ARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
IVASKRKLGETVDITELS[illegible]EQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
IQAILRASRDHDLLVIGA[illegible]SPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
ARRILVPIIGLEYSFAAA[illegible]LAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
ARSVVDEAVFRGRRLGVR[illegible]SSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
YLGSTVESVVVRSRVPVA[illegible]LVAHGGTREQVR"
CDS complement(5612.[illegible] .5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF269[illegible]07.1"
/db_xref="GI:6724241"
/translation="MDKPI[illegible]RTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
QVAITVYFSDRDPAGPYR[illegible]VPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
QHTRLDSRQAENALLSTIA[illegible]TDRE"
CDS 6226. .6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:672424[illegible]"
/translation="MKHVD[illegible]GRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
GADARVTADVDPDAATTRL[illegible]VDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
YNADSRRGKLAETTVPYAN[illegible]ELLITAEKQSSPQSPSSAAVIGPTSVG"
gene 7610. .11875
/gene="epoA"
CDS 7610. .11875
/gene="epoA"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRPI[illegible]RAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
GRVPAERWDAAAWFDPDPDA[illegible]GKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSAL[illegible]GTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
PSVGAGRISYALGLRGPCVA[illegible]DTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALARDGRCKA[illegible]SAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQ[illegible]IVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSV[illegible]TNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWP[illegible]WNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTASALDAQA[illegible]RLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SREGLRAALDAAAQGQTSPG[illegible]VRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
REAFDLCVRLFNQELDRPLR[illegible]VMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAAC[illegible]AGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
DVAAAVAPHAASVSIAAVNA[illegible]DQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYR[illegible]PSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG

VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature 7640. .8920
/gene="epoA"
/note="Region: beta-ketoacyl synthase/malonyl-CoA decarboxylase"
misc_feature 9236. .10201
/gene="epoA"
/note="Region: acetate acyltransferase"
misc_feature 10529. .11428
/gene="epoA"
/note="Region: enoyl reductase"
misc_feature 11549. .11764
/gene="epoA"
/note="Region: acyl carrier protein"
gene 11872. .16104
/gene="epoP"
CDS 11872. .16104
/gene="epoP"
/note="EPOS P; contains condensation/heterocyclization, cysteine adenylation, putative NAD(P)H oxidase, and peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene 16251. .21749
/gene="epoB"
CDS 16251. .21749
/gene="epoB"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 96.8%; Score 97.8; DB 1; Length 68750;

```
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 4
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 5
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 6
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 7
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912
```

```
RESULT 8
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a   22456 c   25539 g   11159 t
ORIGIN

  Query Match             96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||| |||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 9
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a   22456 c   25539 g   11159 t
ORIGIN

  Query Match             96.8%;  Score 97.8;  DB 6;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 2.1e-09;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||| |||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 10
AX024384
LOCUS       AX024384               16124 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024384
VERSION     AX024384.1  GI:10184588
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             62.0%;  Score 62.6;  DB 1;  Length 16124;
  Best Local Similarity   76.2%;  Pred. No. 0.0081;
  Matches   77;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||| ||| | ||||||||| ||||| ||| | ||||| || ||||| ||   |||||||||
Db   1721 CGCCGCGCCGGAGGCCGACGTGGCTGCCGCGGTGGCGCCGCACGCAGCGTTGGTGTCGAT 1780

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||| || ||||| || || ||    || ||||||| ||||
Db   1781 CGCGGCAGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCG 1821


RESULT 11
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             62.0%;  Score 62.6;  DB 6;  Length 16124;
  Best Local Similarity   76.2%;  Pred. No. 0.0081;
  Matches   77;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||| ||| | ||||||||| ||||| ||| | ||||| || ||||| ||   |||||||||
Db   1721 CGCCGCGCCGGAGGCCGACGTGGCTGCCGCGGTGGCGCCGCACGCAGCGTTGGTGTCGAT 1780

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||| || ||||| || || ||    || ||||||| ||||
Db   1781 CGCGGCAGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCG 1821


RESULT 12
SC1G7/c
LOCUS       SC1G7                  26195 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 1G7.
ACCESSION   AL591083 AL645882
VERSION     AL591083.2  GI:20520686
```

KEYWORDS decarboxylase; diaminobutyrate-pyruvate aminotransferase; epoxide hydrolase; FAD-binding protein; integral membrane transport protein; oxidoreductase; regulatory protein; scoT; secreted protein; type I polyketide synthase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 26195)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843436
REFERENCE 2 (bases 1 to 26195)
AUTHORS Seeger,K. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 26195)
AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced gi:13940311.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 1G7 overlaps cosmid St1C4 and cosmid StBAC8D1.

FEATURES Location/Qualifiers

source 1. .26195
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 1G7"

misc_feature 2. .118
/note="nominal overlap with Streptomyces coelicolor cosmid SC2C4"

misc_feature complement(4. .618)
/note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 5.10, E-value 2.8e-10"

misc_feature complement(928. .1437)
/note="Pfam match to entry PF02801 ketoacyl-synt_C, Beta-ketoacyl synthase, score 346.10, E-value 3.8e-100"

misc_feature complement(1459. .2214)
/note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 469.40, E-value 3e-137"

misc_feature complement(1672. .1722)
/note="PS00606 Beta-ketoacyl synthases active site"

misc_feature complement(2284. .2487)
/note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 66.20, E-value 1.1e-16"

misc_feature complement(2362. .2409)
/note="PS00012 Phosphopantetheine attachment site"

misc_feature complement(4861. .5829)
/note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 307.60, E-value 1.4e-88"

misc_feature complement(6142. .6651)
/note="Pfam match to entry PF02801 ketoacyl-synt_C, Beta-ketoacyl synthase, score 328.60, E-value 7.1e-95"

misc_feature complement(6673. .7425)
/note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 458.60, E-value 5.1e-134"

misc_feature complement(7495. .7698)
/note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 55.10, E-value 1.3e-13"

misc_feature complement(7909. .8832)
/note="Pfam match to entry PF00698 Acyl_transf, Acyl transferase domain, score 280.60, E-value 2e-80"

misc_feature complement(9139. .9654)
/note="Pfam match to entry PF02801 ketoacyl-synt_C, Beta-ketoacyl synthase, score 274.60, E-value 1.3e-78"

misc_feature complement(9676. .10416)
/note="Pfam match to entry PF00109 ketoacyl-synt, Beta-ketoacyl synthase, score 353.40, E-value 2.5e-102"

RBS 10776. .10781

gene 10793. .12115
/gene="SC1G7.02"
/note="SCO6276"

CDS 10793. .12115
/gene="SC1G7.02"
/note="SC1G7.02, possible secreted protein, len: 440 aa; similar to TR:Q9L0A8 (EMBL:AL163003) Streptomyces coelicolor putative secreted protein SCC24.09, 468 aa; fasta scores: opt: 777 Z-score: 867.4 bits: 169.7 E(): 1e-40; 36.303% identity in 449 aa overlap. Contains Pfam match to entry PF01360 Monooxygenase, Monooxygenase"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAC37877.1"
/db_xref="GI:13940313"
/db_xref="SPTREMBL:Q93S13"
/translation="MRRKREGRAIVLGGSMAGLFSARALSETFANVTVVDRDELTGPN
ELRRGLPQGHHLHGLLARGQQITEEYFPGITEQMRAAGAQTGDVADDVRWIINGKRLA
KVDSGLIALTSSRPFLERHVRDRVLALPEVSFLERATVQSLVTTADRREVNGVVVRRA
DGTEETLHADLVVDATGRGSRTSVWLQEFGYEAVQEEKHKIGLGYTTRYYRVPDEAFE
GNISINTVASAGVPRGAICQKIDGDRAIVTAYGILGDHPPTDPEGFIGFIKSLAAPDI
HEVLTVSEPIDDPVAYKFPTNLRRHYQRMTDFPAGLLVIGDAVCSFNPSYAQGMTVSA
LGSLVLRRHLADTHRPDPLRYFEDLARDAVDGCWEMAVGADLSFPEVEGERTPEVQEA
HAFIAQVQEAATRDAAVARAYSRVIGLVDSPAVLHDSPIEAAIAGAQR"

misc_feature 11294. .11908
/gene="SC1G7.02"
/note="Pfam match to entry PF01360 Monooxygenase, Monooxygenase, score 4.10, E-value 3.3e-05"

RBS 12101. .12104
/gene="SC1G7.02"

gene 12112. .12972
/gene="SC1G7.03"
/note="SCO6277"

CDS 12112. .12972
/gene="SC1G7.03"
/note="SC1G7.03, possible epoxide hydrolase, len: 286 aa;

```
                     similar to TR:Q4[illegible]566 (EMBL:D16628) Arabidopsis thaliana
                     epoxide hydrolase (EC 3.3.2.3) Seh or F18A8.11, 321 aa;
                     fasta scores: opt: 351 Z-score: 406.4 bits: 83.3 E():
                     4.8e-15; 29.560% identity in 318 aa overlap. Contains Pfam
                     match to entry PF00561 abhydrolase, alpha/beta hydrolase
                     fold"
                     /codon_start=1
                     /transl_table=11
                     /product="putative epoxide hydrolase"
                     /protein_id="CAC37878.1"
                     /db_xref="GI:13940314"
                     /db_xref="SPTREMBL:Q93S12"
                     /translation="MTAIRIGLDGLVFDVDVTGPEDGEPVLLLHGFPHNKESWTETAP
                     LLHAAGLRTVAPDQRGYSPDARPAAVADYRLPHLAADALGVLDALGVSSAHVVGHDWG
                     AAVAWYLAARHPGRVRTLTALAIPHLDAYQHAYRVDEEQQHSSQYVEFLVSDGAADHF
                     LADGAAQLHAWFTQAGEGVLTDAQIARYVATHTAAGTLDAALNWYRANNLLGDPLDFG
                     PVTVPTTFVYSRTDTAVSELAVAKTEEYVTGPYRLLTLEKTSHWQPQQDPDTVAAEIL
                     ARVRAYRQTH"
     misc_feature    12262. .12948
                     /gene="SC1G7.03"
                     /note="Pfam match to entry PF00561 abhydrolase, alpha/beta
                     hydrolase fold, score 74.20, E-value 2.7e-18"
     RBS             13005. .13008
     gene            13015. .14646
                     /gene="SC1G7.04"
                     /note="SCO6278"
     CDS             13015. .14646
                     /gene="SC1G7.04"
                     /note="SC1G7.04, possible integral membrane transport
                     protein, len: 543 aa; similar to TR:Q9RMF9 (EMBL:AF201913)
                     Streptomyces globisporus transmembrane membrane efflux
                     protein SgcB, 521 aa; fasta scores: opt: 1451 Z-score:
                     1413.5 bits: 271.2 E(): 3.8e-71; 46.243% identity in 519
                     aa overlap. Contains possible hydrophobic membrane
                     spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane transport protein"
                     /protein_id="CAC37879.1"
                     /db_xref="GI:13940315"
                     /db_xref="SPTREMBL:Q93S11"
                     /translation="MSSNTAPAVELRAGPREWFGLLALLLPVTLMTADLGVLWLATPY
                     LTADLQPTSSQLLWTTDLYGFMTCGFLVVMGTLGDRIGRRRLLILGSLGVIAASVLAA
                     YSTSPEMLIVARALLGVAGAAVLPSTLSLIIHMFKDDRQRATAIATWVTALSVGIAIG
                     PVIGGVLLEHWWWGSVFLMGVPVMLVPVLLAPVLLPEYKDPGAGRLDLASVVLFLAAI
                     LPVVYGIKKFAEHGWSLANLAAIVIGAAFTVVFVRRQNSLETPLLDMRLFRTRAFTGA
                     LLTLLFGMMALNGVEYLVPQYLLVAGELTPLAAGLWLLPGAAGLIIGSQLTPVLAKRF
                     RPAYVITAGLVVTLIGFWLFAVAGPDDSGIVPAAAGLAVIMFGVAPISVLGTALASGA
                     APPEKAGAASATGQTAYDLGLAFGIAVTGSVAVAVYRSGIADTAPAGIPAAAEEAARD
                     TVGGAAKAAESLPQDVGEQLLTAAREAFTSGFHATAWVSAGMAVLTAVVALVLLRHIP

  Query Match             51.7%;  Score 52.2;  DB 1;  Length 26195;
  Best Local Similarity   71.1%;  Pred. No. 0.57;
  Matches   69;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           || |||||||  ||||||  ||||||  | || ||  ||| || | |||||    ||||||
Db     204 GCCTCCGAGGAGGAGGTGTACGCCTCGATCGCGCCGTACGGCGACGCGGTGGGCATCGCC 145

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           | |||||||||| || ||  |  | |||||| | || |
Db     144 GGGGTCAACGGCCCCGAGTCGTTGGTGATCTCCGGGG 108


RESULT 13
AX211706
LOCUS       AX211706               27541 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 2 from Patent WO0159126.
ACCESSION   AX211706
VERSION     AX211706.1  GI:15523938
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 27541)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 2 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .27541
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     3517 a  10766 c   9529 g   3729 t
ORIGIN

  Query Match             50.1%;  Score 50.6;  DB 6;  Length 27541;
  Best Local Similarity   70.1%;  Pred. No. 1.1;
  Matches   68;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  | |||||   | || ||    || ||| | || || || ||||||| ||||||||
Db  13766 GCGATCCAGGCCACCGAGGACGAGATCGCGGCGCACCTCGACGACACGGTGGCGATCGCC 13825

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||  |  |||| |||||| | || |
Db  13826 GCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTG 13862


RESULT 14
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
```

```
                /function="presumably post-translationally modifies ACP
                domains on PKS"
                /note="putative 4'-phosphopantheteine transferase"
                /codon_start=1
                /transl_table=11
                /product="NysF"
                /protein_id="AAF71762.1"
                /db_xref="GI:8050836"
                /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                PAPTPRRDREHRTHLTVNSPLPRPTFG"
gene            complement(867. .2684)
                /gene="nysG"
CDS             complement(867. .2684)
                /gene="nysG"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysG"
                /protein_id="AAF71763.1"
                /db_xref="GI:8050837"
                /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
gene            complement(2662. .4416)
                /gene="nysH"
CDS             complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
gene            4714. .5748
                /gene="nysDIII"
CDS             4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
gene            5930. .34363
                /gene="nysI"
CDS             5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
```

PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPDALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVERLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA

Query Match 50.1%; Score 50.6; DB 1; Length 123580;
Best Local Similarity 70.1%; Pred. No. 0.75;
Matches 68; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  | |||||   | || ||    || ||| | || || || ||||||| ||||||||
Db   13358 GCGATCCAGGCCACCGAGGACGAGATCGCGGCGCACCTCGACGACACGGTGGCGATCGCC 13417

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||  |  |||| |||||  | || |
Db   13418 GCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTG 13454
```

RESULT 15
AX211739
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS .
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES Location/Qualifiers
source 1. .125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 50.1%; Score 50.6; DB 6; Length 125401;
Best Local Similarity 70.1%; Pred. No. 0.74;
Matches 68; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  | |||||   | || ||    || ||| | || || || ||||||| ||||||||
Db   13766 GCGATCCAGGCCACCGAGGACGAGATCGCGGCGCACCTCGACGACACGGTGGCGATCGCC 13825
```

```
Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||  |  |||| |||||| | || |
Db   13826 GCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTG 13862
```

Search completed: November 5, 2002, 12:29:28
Job time : 625.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_27200_27300
Perfect score: 101
Sequence: 1 cgcagcgtccgaggccgagg.....cgccgtcgtgatcgctggcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	71989	21	AAA29349	Sorangium cellulos
2	97.8	96.8	68750	21	AAZ55887	Sorangium cellulos
3	50.6	50.1	27541	22	AAD17185	Streptomyces nours
4	50.6	50.1	125401	22	AAD17186	Streptomyces nours
5	47.4	46.9	15872	18	AAT68715	Streptomyces venez
6	47.4	46.9	15872	21	AAZ87283	S. venezuelae vep
7	45.6	45.1	53789	19	AAV21187	Amycolatopsis medi
8	44.2	43.8	3978	21	AAC55785	Type I polyketide
9	44.2	43.8	12249	21	AAC55840	Complete Mitomycin
10	44.2	43.8	18331	21	AAC55857	Complete nucleotid
11	43.4	43.0	5544	22	AAF24891	Pimaricin biosynth
12	41.2	40.8	29879	14	AAQ46806	eryA region of S.
13	41	40.6	1030	20	AAV99230	DNA encoding an ac
14	41	40.6	1030	21	AAA72514	Nucleotide sequenc
15	41	40.6	44377	18	AAT78508	Platenolide syntha
16	41	40.6	44377	18	AAT80414	Platenolide syntha
17	41	40.6	65140	22	AAD17184	Streptomyces nours
18	39.4	39.0	43280	18	AAT80413	Tylactone synthase
c 19	38.6	38.2	1028	13	AAQ27091	XTY26 probe. Homo
c 20	38.6	38.2	1028	22	AAA91542	1.0kb PstI fragmen
c 21	38.6	38.2	1028	22	AAS01347	Human Fragile X Sy
22	37.8	37.4	31422	22	AAH79278	Streptomyces averm
23	37.6	37.2	20394	22	AAF24892	Pimaricin biosynth
24	37.6	37.2	47981	22	AAF30757	Micromonospora meg
c 25	37	36.6	3765	24	ABK84072	Human cDNA differe
c 26	37	36.6	4362	24	ABK84641	Human cDNA differe
27	36.2	35.8	31422	21	AAA92302	S. avermitilis ave
c 28	35.6	35.2	546	24	ABQ90865	M. capsulatus gene
c 29	35.4	35.0	567	21	AAA29550	HIV codon altered
c 30	35.4	35.0	3765	13	AAQ31890	FMR-1 gene for det
31	35.4	35.0	37856	21	AAA11992	S. cellulosum DNA
32	35	34.7	1014	23	ABL25521	Drosophila melanog
33	35	34.7	3596	23	ABL25520	Drosophila melanog
34	35	34.7	23666	12	AAQ10190	Cephalosporin anti
35	34.6	34.3	1871	24	ABK11624	S. erythraea monen
c 36	34.6	34.3	5746	22	AAH26230	Neurospora crassa
37	34.6	34.3	50937	21	AAA09469	Streptococcus olea
38	34.4	34.1	14775	22	AAF88338	S. spinosa DNA fra
39	34.4	34.1	50000	22	AAF88313	S. spinosa DNA fra
40	34.4	34.1	50000	22	AAF88316	S. spinosa DNA fra
41	34.4	34.1	80161	20	AAZ21501	DNA fragment of Sa
42	34	33.7	2539	18	AAT61454	Streptomyces virid
43	34	33.7	2809	18	AAT61455	DhpA-mel chimeric
44	34	33.7	28598	17	AAT06769	Sorangium cellulos
45	34	33.7	28958	18	AAT89956	Sorangium cellulos

ALIGNMENTS

RESULT 1
AAA29349
ID AAA29349 standard; DNA; 71989 BP.
XX
AC AAA29349;
XX
DT 12-SEP-2000 (first entry)
XX
DE Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /*tag= c
FT /label= epoA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548

```
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                        domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                        domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                        domain, potentially involved in formation of the
FT                        thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                        domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                        module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                        module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                        module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                        NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                        module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                        module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                        module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                        module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 2e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  27200 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 27259

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27260 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 27300


RESULT 2
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
```

```
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR          AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR          AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR          AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             96.8%;  Score 97.8;  DB 21;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 1.1e-14;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||| ||||
Db   32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 3
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             50.1%;  Score 50.6;  DB 22;  Length 27541;
  Best Local Similarity   70.1%;  Pred. No. 0.0014;
  Matches   68;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  | |||||  | || ||    || ||| | || || || ||||||| ||||||||
Db   13766 GCGATCCAGGCCACCGAGGACGAGATCGCGGCGCACCTCGACGACACGGTGGCGATCGCC 13825

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||  |  |||| |||||| | || |
Db   13826 GCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTG 13862


RESULT 4
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
```

```
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match            50.1%;  Score 50.6;  DB 22;  Length 125401;
  Best Local Similarity   70.1%;  Pred. No. 0.0013;
  Matches   68;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
            |||  | |||||   | || ||    || ||| | || || || ||||||| ||||||||
Db   13766 GCGATCCAGGCCACCGAGGACGAGATCGCGGCGCACCTCGACGACACGGTGGCGATCGCC 13825

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||  |  |||| |||||| | || |
Db   13826 GCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTG 13862


RESULT 5
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT   CDS             14056..14136
FT                   /*tag= b
FT   CDS             14148..15827
FT                   /*tag= c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
```

```
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             46.9%;  Score 47.4;  DB 18;  Length 15872;
  Best Local Similarity   68.0%;  Pred. No. 0.0082;
  Matches   66;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  |||||||||||| | ||  ||| ||     | |||. | || |||  |  |||||
Db    5150 GCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTCGCC 5209

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||   |||||| ||  || | ||||
Db    5210 GCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCG 5246


RESULT 6
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag= b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag= c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Désosamine and macrolide biosynthetic gene clusters, useful for, e.g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match             46.9%;  Score 47.4;  DB 21;  Length 15872;
  Best Local Similarity   68.0%;  Pred. No. 0.0082;
  Matches   66;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  |||||||||||| | ||  ||| ||     | |||  | || |||  |  |||||
Db    5150 GCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTCGCC 5209

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| ||   |||||| ||  || | ||||
Db    5210 GCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCG 5246


RESULT 7
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
```

```
FT   CDS             41373..51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 19;  Length 53789;
  Best Local Similarity   68.5%;  Pred. No. 0.02;
  Matches   63;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      10 CGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCGGT 69
           | | | |||||||||| |||    | || ||    || | |   ||   ||||||||| ||
Db   22698 CCACGACGAGGTGGCTGCCCGGCTCGCGCCGTGGGCGGGCCGCGTCGAGATCGCCGCCGT 22757

Qy      70 CAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           ||||||||| | | | ||||||||||| || |
Db   22758 CAACGGTCCGGCCTCGGTCGTGATCGCCGGTG 22789


RESULT 8
AAC55785
ID   AAC55785 standard; DNA; 3978 BP.
XX
AC   AAC55785;
XX
DT   19-JAN-2001  (first entry)
XX
DE   Type I polyketide synthase orf8.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999;   99US-0266965.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH,  Mao Y,  Varoglu M,  He M,  Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis -
XX
PS   Example 1; Page 251-253; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 3978 BP; 497 A; 1583 C; 1415 G; 483 T; 0 other;

  Query Match             43.8%;  Score 44.2;  DB 21;  Length 3978;
  Best Local Similarity   66.0%;  Pred. No. 0.05;
  Matches   64;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           ||| |||||| |||||||  ||||    | ||   |  |  | ||   |||   ||||||
Db    2122 GCGACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCC 2181

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| || || || || ||| |  | ||||
Db    2182 GCCGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCG 2218


RESULT 9
```

```
AAC55840
ID   AAC55840 standard; DNA; 12249 BP.
XX
AC   AAC55840;
XX
DT   19-JAN-2001  (first entry)
XX
DE   Complete Mitomycin ORF 1-9 nucleotide sequence.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999;   99US-0266965.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH,  Mao Y,  Varoglu M,  He M,  Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis -
XX
PS   Claim 26; Figure 22; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 12249 BP; 1715 A; 4722 C; 4196 G; 1616 T; 0 other;

  Query Match           43.8%;  Score 44.2;  DB 21;  Length 12249;
  Best Local Similarity  66.0%;  Pred. No. 0.047;
  Matches   64;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           ||| |||||| ||||||| |||| | ||  |  |  | ||  |||  ||||||
Db    3109 GCGACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCC 3168

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| || || || || ||| |  | ||||
Db    3169 GCCGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCG 3205


RESULT 10
AAC55857
ID   AAC55857 standard; DNA; 18331 BP.
XX
AC   AAC55857;
XX
DT   19-JAN-2001  (first entry)
XX
DE   Complete nucleotide sequence of the mitomycin gene cluster.
XX
KW   Mitomycin; biosynthesis; mitosane ring system; antibiotic; anti-cancer;
KW   anti-inflammatory; immune-enhancer; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   fungicide; pesticide; ds.
XX
OS   Streptomyces lavendulae.
XX
PN   WO200053737-A2.
XX
PD   14-SEP-2000.
XX
PF   10-MAR-2000; 2000WO-US06394.
XX
PR   12-MAR-1999;   99US-0266965.
XX
PA   (MINU ) UNIV MINNESOTA.
PA   (SHER/) SHERMAN D H.
PA   (MAOY/) MAO Y.
PA   (VARO/) VAROGLU M.
PA   (HEMM/) HE M.
PA   (SHEL/) SHELDON P C.
XX
PI   Sherman DH,  Mao Y,  Varoglu M,  He M,  Sheldon PC;
XX
DR   WPI; 2000-601980/57.
XX
PT   Novel nucleic acid molecule comprising mitomycin biosynthetic gene
PT   cluster useful for cloning mitomycin biosynthetic genes for elucidating
PT   the molecular basis of mitosane ring system biosynthesis -
XX
PS   Disclosure; Figure 21; 399pp; English.
XX
CC   This invention relates to isolated and purified nucleic acid molecules
CC   from the mitomycin biosynthetic gene cluster. Mitomycins are a group of
CC   natural products that contain a variety of functional groups, including
CC   amino benzoquinone and axiridine ring systems. The S. lavendulae
CC   mitomycin biosynthetic gene cluster comprises 47 mitomycin genes
CC   spanning 55kb of DNA. The invention includes an expression cassette
CC   comprising a mitomycin biosynthetic gene operably linked to a promoter,
CC   and host cells transformed with the cassette. The nucleotide, and protein
CC   sequences and the transformed host cells of the invention result in
CC   antiasthmatic, antiinflammatory, cytostatic, immunomodulatory, and
CC   antibiotic activities. The nucleotide sequences are used to elucidate the
CC   molecular basis for the biosynthesis of the mitosane ring system, as well
CC   as to engineer the biosynthesis of novel natural products, e.g.
CC   antibiotics, anti-inflammatory agents, anti-cancer agents,
CC   immune-enhancers, immunosuppressants, agents to treat asthma, chronic
CC   obstructive pulmonary disease as well as other disease involving
CC   respiratory inflammation, or cholesterol-lowering agents or as crop
CC   protection agents (e.g. fungicides or insecticides) as well as
CC   bipolymers, e.g., in packaging or biomedical applications, or to engineer
CC   PHA monomer synthases. Sequences AAC55782-C55881, AAC55815-C55849 and
CC   AAB32485-B32542 represent mitomycin biosynthetic gene cluster DNA
CC   sequences and encoded proteins. Sequences AAC55812-C55814,
CC   AAC55850-C55856 and AAC55862-C55869 represent PCR primers used in the
```

```
CC   cloning of the mitomycin biosynthetic genes.
XX
SQ   Sequence 18331 BP; 2523 A; 7003 C; 6343 G; 2462 T; 0 other;

  Query Match             43.8%;  Score 44.2;  DB 21;  Length 18331;
  Best Local Similarity   66.0%;  Pred. No. 0.046;
  Matches   64;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           ||| |||||| |||||||  ||||    | ||   |  |  | ||   |||   ||||||
Db    3109 GCGACCGAGGACGAGGTGCGCGCCCAACTCGCCGACGGCCGCCCCGGCGTGGACATCGCC 3168

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || |||||||| || || || || ||| |  | ||||
Db    3169 GCCGTCAACGGACCGGAAGCGGTGGTGCTGTCCGGCG 3205


RESULT 11
AAF24891
ID   AAF24891 standard; DNA; 5544 BP.
XX
AC   AAF24891;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5544
FT                   /*tag=  a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31557.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 39-46; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 5544 BP; 699 A; 2052 C; 2023 G; 770 T; 0 other;

  Query Match             43.0%;  Score 43.4;  DB 22;  Length 5544;
  Best Local Similarity   66.7%;  Pred. No. 0.076;
  Matches   62;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       9 CCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCGG 68
           | ||  |||| ||| || | |   | ||        | | |  ||||||||||| || |
Db    4037 CGGAACCCGAAGTGACCCCATGGCTCGCGGAGTTGACGGACGAGGTGTCGATCGCGGCCG 4096

Qy      69 TCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           ||||||||||  || || ||||| |||| ||||
Db    4097 TCAACGGTCCGCACTCCCTCGTGCTCGCGGGCG 4129


RESULT 12
AAQ46806
ID   AAQ46806 standard; DNA; 29879 BP.
XX
AC   AAQ46806;
XX
DT   22-DEC-1993  (first entry)
XX
DE   eryA region of S. erythraea chromosome.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement; ss.
XX
OS   Saccarapolyspora erythraea.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..10199
FT                   /*tag=  a
FT                   /note= "ORF 1"
FT   CDS             10218..20921
FT                   /*tag=  b
FT                   /note= "ORF 2"
FT   CDS             20922..29879
FT                   /*tag=  c
FT                   /note= "ORF 3"
XX
PN   WO9313663-A.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US00427.
XX
PR   17-JAN-1992;   92WO-US00427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Donadio S,  Katz L,  McAlpine JB;
XX
DR   WPI; 1993-242804/30.
DR   P-PSDB; AAR44430-32.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin
PT   cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT   microorganisms
XX
PS   Claim 27; Fig 2; 133pp; English.
XX
CC   This sequence represents a fragment of the Saccarapolyspora erythraea
CC   genome, designated eryA.  The polypeptides encoded by this region
CC   are involved in the biosynthesis of the polyketide segment of
CC   erythromycin.  eryA is organised in modules and each module takes care
CC   of one condensation step.  The precise succession of elongation steps
CC   is dictated by the genetic order of the modules.  This fragment may be
CC   specifically altered such that novel polyketide molecules of desired
CC   structure are produced.  Three types of alteration may be produced;
CC   those inactivating a single function in a module which does not arrest
CC   acyl chain growth; those inactivating a single function in a module
CC   which does affect chain growth; and those affecting an entire module.
CC   The mutations may be introduced by gene replacement.
XX
SQ   Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;
```

```
  Query Match           40.8%;  Score 41.2;  DB 14;  Length 29879;
  Best Local Similarity  64.9%;  Pred. No. 0.22;
  Matches   61;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       8 TCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCG 67
           ||||| | ||||  |     | |||  |     |    ||| || | ||  | ||||||||
Db    7653 TCCGACGTCGAGAAGCTGCTCGCCGACGACAGCTGGACCGGCAGGCTGGAGGTCGCCGCG 7712

Qy      68 GTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||| || |||||||| ||| | || || |
Db    7713 GTCAACGGCCCCGACGCCGTGGTGGTGGCCGGTG 7746


RESULT 13
AAV99230
ID   AAV99230 standard; DNA; 1030 BP.
XX
AC   AAV99230;
XX
DT   09-MAR-1999  (first entry)
XX
DE   DNA encoding an active acyltransferase domain (venAT) from a PKS.
XX
KW   Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW   polyketide synthesis; antimicrobial; anticancer; antifungal;
KW   immunosuppressant; anthelmintic; venAT; malonate acyltransferase domain;
KW   PKS cluster; Streptomyces venezuelae ATCC15439; ds.
XX
OS   Streptomyces venezuelae.
XX
PN   WO9851695-A2.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-US09518.
XX
PR   16-MAY-1997;   97US-0858003.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Kakavas SJ,  Katz L,  Pereda-Lopez A,  Ruan X,  Stassi DL;
PI   Summers RG;
XX
DR   WPI; 1999-070114/06.
DR   P-PSDB; AAW87713.
XX
PT   New polyketides produced by microorganism having altered polyketide
PT   synthase genes - especially chimeras containing a heterologous
PT   acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT   agents
XX
PS   Claim 24; Fig 13; 137pp; English.
XX
CC   The present sequence encodes an enzymatically active acyltransferase
CC   domain, venAT, from a polyketide synthase (PKS). The venAT protein
CC   is derived from the malonate acyltransferase domain from the
CC   PKS cluster of Streptomyces venezuelae ATCC15439. The sequence
CC   can be used to replace acyltransferase domains of other PKS enzymes,
CC   which are then used to generate novel polyketides by de novo biosynthesis
CC   rather than by chemical modification. Polyketides of the invention have
CC   the methyl groups on the macrolactone ring substituted with -H, -Et
CC   and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC   dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC   antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC   anthelmintic agents, or their intermediates.
XX
SQ   Sequence 1030 BP; 132 A; 383 C; 362 G; 153 T; 0 other;

  Query Match           40.6%;  Score 41;  DB 20;  Length 1030;
  Best Local Similarity  63.9%;  Pred. No. 0.31;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  ||||| ||||||   ||   || | ||  ||  || ||   ||||  |  | |||
Db     421 GCGAGCGAGGACGAGGTACGCGAGGCCCTCGCGACCTTCGACGATGCGGTTGCCGTGGCC 480

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |  |||||||| ||  ||||  |||  || | ||||
Db     481 GGAGTCAACGGCCCGAACGCCACCGTCGTCTCCGGCG 517


RESULT 14
AAA72514
ID   AAA72514 standard; DNA; 1030 BP.
XX
AC   AAA72514;
XX
DT   20-NOV-2000  (first entry)
XX
DE   Nucleotide sequence of the venAT domain.
XX
KW   Erythromycin derivative; polyketide synthase; PKS; acetyltransferase;
KW   antibiotic; antifungal; anticancer; immunosuppressant; antihelminithic;
KW   venAT; ds.
XX
OS   Streptomyces venezuelae.
XX
PN   US6063561-A.
XX
PD   16-MAY-2000.
XX
PF   13-MAY-1998;   98US-0078166.
XX
PR   17-JAN-1991;   91US-0642734.
PR   16-MAY-1997;   97US-0858003.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Katz L,  Stassi DL,  Ruan X,  Pereda-Lopez A,  Kakavas SJ,  Summers RG;
XX
DR   WPI; 2000-386760/33.
DR   P-PSDB; AAB12925.
XX
PT   New erythromycin derivatives produced by microorganisms with altered
PT   polyketide synthase genes  -
XX
PS   Example 9; Fig 13; 94pp; English.
XX
CC   This invention relates to novel erythromycin derivatives, in which the
CC   methyl groups on the macrolactone ring have been substituted  with H, Et
CC   and/or OH and the ethyl side chain has been substituted with
CC   hydroxymethyl or dihydroxycyclohexylmethyl side chains. Included in the
CC   invention are vectors containing polynucleotide sequences which encode an
CC   enzymatically active acetyltransferase domain from Streptomyces. Also
CC   included are methods for altering the substrate specificity of a
CC   polyketide synthase in a microorganism using the vectors of the
CC   invention. Polyketides are are a large class of natural products
CC   including antibiotics, antifungal, anticancer and immunosuppressant
CC   compounds. Polyketide synthase (PKS) is a complex of enzyme activities
CC   responsible for the biosynthesis of polyketides. The erythromycin
CC   derivatives exhibit antibiotic, antifungal, anticancer, antihelminithic
CC   and immunosuppressant activity.
CC   The present sequence represents the nucleotide sequence encoding the
CC   venAT domain of Streptomyces venezuelae. The sequence is used in the
CC   construction of the a vector of the invention.
XX
SQ   Sequence 1030 BP; 132 A; 383 C; 362 G; 153 T; 0 other;

  Query Match           40.6%;  Score 41;  DB 21;  Length 1030;
  Best Local Similarity  63.9%;  Pred. No. 0.31;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  ||||| ||||||   ||   || | ||  ||  || ||   ||||  |  | |||
Db     421 GCGAGCGAGGACGAGGTACGCGAGGCCCTCGCGACCTTCGACGATGCGGTTGCCGTGGCC 480
```

```
Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |  |||||||| ||  |||||  |||  || | ||||
Db     481 GGAGTCAACGGCCCGAACGCCACCGTCGTCTCCGGCG 517


RESULT 15
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             40.6%;  Score 41;  DB 18;  Length 44377;
  Best Local Similarity   63.9%;  Pred. No. 0.24;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           ||  |||||   ||||  |||||    | | |   |  ||||| |   | |  | |||||
Db   38309 GCCACCGAGCGGGAGGCCGCCGCGCTGGAGCCCGTCGCCGCCGGCGGCGCGGTGGTCGCC 38368

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           ||||||||||| ||  | ||| ||||| || | || |
Db   38369 GCGGTCAACGGCCCGCAGGCCCTCGTGCTCTCCGGGG 38405


Search completed: November  5, 2002, 13:38:07
Job time : 223.215 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_27200_27300
Perfect score: 101
Sequence: 1 cgcagcgtccgaggccgagg.....cgccgtcgtgatcgctggcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	2	97.8	96.8	68750	3	US-09-335-409-1	Sequence 1, Appli
	3	97.8	96.8	68750	4	US-09-568-102-1	Sequence 1, Appli
	4	97.8	96.8	68750	4	US-09-567-969-1	Sequence 1, Appli
	5	97.8	96.8	68750	4	US-09-568-480-1	Sequence 1, Appli
	6	97.8	96.8	68750	4	US-09-568-486-1	Sequence 1, Appli
	7	97.8	96.8	68750	4	US-09-568-472-1	Sequence 1, Appli
	8	97.8	96.8	68750	4	US-09-567-899-1	Sequence 1, Appli
	9	47.4	46.9	15872	4	US-09-105-537-1	Sequence 1, Appli
	10	41.2	40.8	11219	1	US-07-642-734C-1	Sequence 1, Appli
	11	41.2	40.8	11219	3	US-08-439-009A-1	Sequence 1, Appli
	12	41	40.6	1030	3	US-08-858-003-2	Sequence 2, Appli
	13	41	40.6	1030	3	US-09-078-166-2	Sequence 2, Appli
	14	41	40.6	1030	4	US-08-997-467-2	Sequence 2, Appli
	15	41	40.6	44377	2	US-08-804-227C-7	Sequence 7, Appli
	16	41	40.6	44377	2	US-08-804-198-1	Sequence 1, Appli
	17	39.4	39.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	18	38.6	38.2	1028	4	US-08-118-200-1	Sequence 1, Appli
c	19	38.6	38.2	1028	4	US-08-458-745-1	Sequence 1, Appli
	20	38.6	38.2	20235	1	US-07-642-734C-3	Sequence 3, Appli
	21	38.6	38.2	20235	3	US-08-439-009A-3	Sequence 3, Appli
c	22	37	36.6	3765	3	US-07-705-490-1	Sequence 1, Appli
c	23	37	36.6	3765	4	US-07-751-891B-1	Sequence 1, Appli
c	24	37	36.6	4362	2	US-08-455-073A-1	Sequence 1, Appli
c	25	34.6	34.3	90	4	US-07-827-691A-4	Sequence 4, Appli
	26	34.6	34.3	50937	4	US-09-428-517-1	Sequence 1, Appli
	27	34.4	34.1	80161	3	US-09-036-987A-1	Sequence 1, Appli
	28	34.4	34.1	80161	4	US-09-370-700-1	Sequence 1, Appli
	29	34	33.7	2539	3	US-09-000-016-3	Sequence 3, Appli
	30	34	33.7	2539	4	US-09-514-340-3	Sequence 3, Appli
	31	34	33.7	2809	3	US-09-000-016-1	Sequence 1, Appli
	32	34	33.7	2809	4	US-09-514-340-1	Sequence 1, Appli
	33	34	33.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
	34	34	33.7	28958	1	US-08-456-837-6	Sequence 6, Appli
	35	34	33.7	28958	1	US-08-457-342-6	Sequence 6, Appli
	36	34	33.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
	37	34	33.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
	38	34	33.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
	39	34	33.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
	40	34	33.7	28958	1	US-08-729-214-6	Sequence 6, Appli
	41	34	33.7	28958	3	US-09-028-934-6	Sequence 6, Appli
	42	34	33.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	43	34	33.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	44	33.4	33.1	925	3	US-08-858-003-1	Sequence 1, Appli
	45	33.4	33.1	925	3	US-09-078-166-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 7.4e-16;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27200 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 27259

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||||||||||
Db   27260 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 27300
```

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             96.8%;  Score 97.8;  DB 3;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||| ||||
Db   32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||| ||||
Db   32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 4
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||| ||||
Db   32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||||||||||||||||||||||| ||||
Db   32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             96.8%;  Score 97.8;  DB 4;  Length 68750;
  Best Local Similarity   98.0%;  Pred. No. 4.1e-15;
  Matches   99;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  32812 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCGCCCCACGCCGCCACGGTGTCGAT 32871

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||||||||||||||||||||||||||||||||| ||||
Db  32872 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCCGGCG 32912


RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             46.9%;  Score 47.4;  DB 4;  Length 15872;
  Best Local Similarity   68.0%;  Pred. No. 0.0019;
  Matches   66;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
          |||  ||||||||||||| | ||  ||| ||     | |||  | || |||  |  |||||
Db   5150 GCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTCGCC 5209

Qy     65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          || |||||||| ||   |||||| ||  || | ||||
```

```
Db    5210 GCCGTCAACGGACCCACCGCCGTGGTC3TCTCCGGCG 5246


RESULT 10
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1

  Query Match             40.8%;  Score 41.2;  DB 1;  Length 11219;
  Best Local Similarity   64.9%;  Pred. No. 0.053;
  Matches   61;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      8 TCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCG 67
          ||||| | |||| |     | ||| |    |    ||| || | ||  | ||||||||
Db   8673 TCCGACGTCGAGAAGCTGCTCGCCGACGACAGCTGGACCGGCAGGCTGGAGGTCGCCGCG 8732

Qy     68 GTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||| || |||||||| ||| | || || |
Db   8733 GTCAACGGCCCCGACGCCGTGGTGGTGGCCGGTG 8766


RESULT 11
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
```

```
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1

  Query Match             40.8%;  Score 41.2;  DB 3;  Length 11219;
  Best Local Similarity   64.9%;  Pred. No. 0.053;
  Matches   61;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       8 TCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCG 67
           ||||| | ||||  |      | |||  |    |    ||| || | ||  | ||||||||
Db    8673 TCCGACGTCGAGAAGCTGCTCGCCGACGACAGCTGGACCGGCAGGCTGGAGGTCGCCGCG 8732

Qy      68 GTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||| || |||||||| ||| | || || |
Db    8733 GTCAACGGCCCCGACGCCGTGGTGGTGGCCGGTG 8766


RESULT 12
US-08-858-003-2
; Sequence 2, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
```

```
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dianne Casuto
;       REGISTRATION NUMBER:  P-40,943
;       REFERENCE/DOCKET NUMBER:  4952.US.P2
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (847)-938-3137
;       TELEFAX:  (847)-938-2623
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1030 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-858-003-2

  Query Match             40.6%;  Score 41;  DB 3;  Length 1030;
  Best Local Similarity   63.9%;  Pred. No. 0.067;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  ||||| ||||||   ||   || | ||  ||  || ||   |||| |  | |||
Db     421 GCGAGCGAGGACGAGGTACGCGAGGCCCTCGCGACCTTCGACGATGCGGTTGCCGTGGCC 480

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |  |||||||| ||  ||||| |||  || | ||||
Db     481 GGAGTCAACGGCCCGAACGCCACCGTCGTCTCCGGCG 517


RESULT 13
US-09-078-166-2
; Sequence 2, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1030 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-09-078-166-2

  Query Match             40.6%;  Score 41;  DB 3;  Length 1030;
  Best Local Similarity   63.9%;  Pred. No. 0.067;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  ||||| ||||||   ||   || | ||  ||  || ||   |||| |  | |||
Db     421 GCGAGCGAGGACGAGGTACGCGAGGCCCTCGCGACCTTCGACGATGCGGTTGCCGTGGCC 480

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |  |||||||| ||  ||||| |||  || | ||||
Db     481 GGAGTCAACGGCCCGAACGCCACCGTCGTCTCCGGCG 517


RESULT 14
US-08-997-467-2
; Sequence 2, Application US/08997467
; Patent No. 6200813
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,467
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;       LENGTH:  1030 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
US-08-997-467-2

  Query Match             40.6%;  Score 41;  DB 4;  Length 1030;
  Best Local Similarity   63.9%;  Pred. No. 0.067;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
```

```
          ||| ||||| ||||||   ||   || | ||  ||  || ||   |||| |  | |||
Db    421 GCGAGCGAGGACGAGGTACGCGAGGCCCTCGCGACCTTCGACGATGCGGTTGCCGTGGCC 480

Qy     65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |  |||||||| ||  |||||  |||  || | ||||
Db    481 GGAGTCAACGGCCCGAACGCCACCGTCGTCTCCGGCG 517


RESULT 15
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             40.6%;  Score 41;  DB 2;  Length 44377;
  Best Local Similarity   63.9%;  Pred. No. 0.055;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
          ||  |||||   ||||  |||||    | | |   |  ||||| |   | |  | |||||
Db  38309 GCCACCGAGCGGGAGGCCGCCGCGCTGGAGCCCGTCGCCGCCGGCGGCGCGGTGGTCGCC 38368

Qy     65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          |||||||||| ||  | ||| ||||| || | || |
Db  38369 GCGGTCAACGGCCCGCAGGCCCTCGTGCTCTCCGGGG 38405


Search completed: November  5, 2002, 15:41:45
Job time : 151.737 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_27200_27300
Perfect score: 101
Sequence: 1 cgcagcgtccgaggccgagg.....cgccgtcgtgatcgctggcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	47.4	46.9	15872	10	US-09-861-289-1	Sequence 1, Appli
	2	33.8	33.5	546	10	US-09-815-242-4036	Sequence 4036, Ap
	3	32.8	32.5	1294	10	US-09-850-964-4	Sequence 4, Appli
	4	32.8	32.5	2289	10	US-09-850-964-3	Sequence 3, Appli
	5	32.2	31.9	1530	10	US-09-815-242-7987	Sequence 7987, Ap
	6	32	31.7	1395	10	US-09-815-242-7769	Sequence 7769, Ap
	7	32	31.7	4041	10	US-09-861-289-36	Sequence 36, Appl
	8	32	31.7	36778	10	US-09-861-289-5	Sequence 5, Appli
	9	31.8	31.5	2724	12	US-10-044-090-309	Sequence 309, App
c	10	31.8	31.5	11188	10	US-09-821-167-1	Sequence 1, Appli
c	11	31.2	30.9	2038	10	US-09-880-107-2102	Sequence 2102, Ap
c	12	31.2	30.9	4689	10	US-09-861-289-34	Sequence 34, Appl
c	13	31.2	30.9	6483	10	US-09-880-107-3699	Sequence 3699, Ap
c	14	31.2	30.9	13842	10	US-09-861-289-30	Sequence 30, Appl
c	15	31.2	30.9	36778	10	US-09-861-289-5	Sequence 5, Appli
	16	31	30.7	251	9	US-09-933-797-319	Sequence 319, App
	17	30.8	30.5	280	10	US-09-294-093B-2177	Sequence 2177, Ap
c	18	30.6	30.3	967	10	US-09-073-009-11	Sequence 11, Appl
c	19	30.6	30.3	967	10	US-09-073-009-83	Sequence 83, Appl
c	20	30.6	30.3	967	10	US-09-023-588-11	Sequence 11, Appl
c	21	30.6	30.3	967	10	US-09-793-306-11	Sequence 11, Appl
c	22	30.6	30.3	967	10	US-09-793-306-83	Sequence 83, Appl
c	23	30.6	30.3	1231	10	US-09-073-009-104	Sequence 104, App
c	24	30.6	30.3	1231	10	US-09-793-306-104	Sequence 104, App
	25	30.6	30.3	1266	10	US-09-815-242-7920	Sequence 7920, Ap
c	26	30.6	30.3	1752	10	US-09-073-009-125	Sequence 125, App
c	27	30.6	30.3	1752	10	US-09-793-306-125	Sequence 125, App
c	28	30.6	30.3	1926	10	US-09-925-301-489	Sequence 489, App
	29	30.4	30.1	1428	10	US-09-815-242-4000	Sequence 4000, Ap
c	30	30.4	30.1	8588	10	US-09-816-669A-13	Sequence 13, Appl
	31	30.4	30.1	11220	10	US-09-861-289-32	Sequence 32, Appl
	32	30.2	29.9	13842	10	US-09-861-289-30	Sequence 30, Appl
c	33	30	29.7	505	10	US-09-867-550-2043	Sequence 2043, Ap
	34	30	29.7	1434	10	US-09-815-242-7729	Sequence 7729, Ap
c	35	29.8	29.5	33023	10	US-09-880-107-3350	Sequence 3350, Ap
	36	29.6	29.3	293	10	US-09-294-093B-3067	Sequence 3067, Ap
	37	29.6	29.3	1294	10	US-09-748-033-2	Sequence 2, Appli
	38	29.4	29.1	2419	10	US-09-939-408A-12	Sequence 12, Appl
	39	29.4	29.1	4039	10	US-09-205-448-7	Sequence 7, Appli
	40	29.2	28.9	218	10	US-09-878-574-123	Sequence 123, App
	41	29.2	28.9	1630	10	US-09-823-038A-4	Sequence 4, Appli
	42	29	28.7	1038	10	US-09-815-242-4048	Sequence 4048, Ap
	43	29	28.7	2178	10	US-09-815-242-4070	Sequence 4070, Ap
	44	28.8	28.5	2314	10	US-09-962-832-105	Sequence 105, App
	45	28.6	28.3	295	10	US-09-294-093B-1139	Sequence 1139, Ap

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 46.9%; Score 47.4; DB 10; Length 15872;
Best Local Similarity 68.0%; Pred. No. 0.00026;
Matches 66; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           |||  |||||||||||| | ||  ||| ||     | |||  | || |||  |  |||||
Db    5150 GCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCCGTCGCC 5209

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             || |||||||| ||   |||||| ||  || | ||||
Db    5210 GCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCG 5246
```

RESULT 2
US-09-815-242-4036
; Sequence 4036, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

```
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4036
;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4036

  Query Match             33.5%;  Score 33.8;  DB 10;  Length 546;
  Best Local Similarity   58.4%;  Pred. No. 0.77;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
             |||  || ||  ||||||||| | || |    | | || |   |  ||| | | | ||||
Db      75 CGCTTCGACCCTGGCCGAGGTCGGCGTCCGTTTCGAACGCTGGGAAGCCGCCGCGCCGAT 134

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||| | |  | | |  |||| || |  |||  ||
Db     135 CGCCGCCGGCGCCAGCCAGGACGAGGTGATCGCCGCCTACG 175


RESULT 3
US-09-850-964-4
; Sequence 4, Application US/09850964
; Patent No. US20020104121A1
; GENERAL INFORMATION:
;  APPLICANT: MCELROY, DAVID
;  APPLICANT:  KRIZ, ALAN L.
;  APPLICANT:  OROZCO, EMIL
;  APPLICANT:  GRIFFOR, MATT
;  TITLE OF INVENTION: THE MAIZE A3 PROMOTER AND RICE ACTIN 2 INTRON AND
;  TITLE OF INVENTION:  METHODS FOR USE THEREOF
;  FILE REFERENCE: DEKM:166
;  CURRENT APPLICATION NUMBER: US/09/850,964
;  CURRENT FILING DATE:  2001-05-07
;  PRIOR APPLICATION NUMBER: 09/312,038
;  PRIOR FILING DATE: 1999-05-14
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 1294
;   TYPE: DNA
;   ORGANISM: Zea mays/rice
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (883)..(972)
;   OTHER INFORMATION: N = A or C or G or T
US-09-850-964-4

  Query Match             32.5%;  Score 32.8;  DB 10;  Length 1294;
  Best Local Similarity   58.0%;  Pred. No. 1.3;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
             || |   |||  ||| ||| ||| || |||| ||  ||  | || |||| ||  | | |
Db     539 GCCGTAGCCGCCGCCACGGTTGCCACCGCCGTAGCCGCCACCACCACCACCGTAGCCACC 598

Qy      62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||   |   |   | | ||||  ||   | || | ||
Db     599 GCCGCCACCGTAGCCGCCGCCGCCACCGCCGTAGCCGCCG 638


RESULT 4
US-09-850-964-3
; Sequence 3, Application US/09850964
; Patent No. US20020104121A1
; GENERAL INFORMATION:
;  APPLICANT: MCELROY, DAVID
;  APPLICANT:  KRIZ, ALAN L.
;  APPLICANT:  OROZCO, EMIL
;  APPLICANT:  GRIFFOR, MATT
;  TITLE OF INVENTION: THE MAIZE A3 PROMOTER AND RICE ACTIN 2 INTRON AND
;  TITLE OF INVENTION:  METHODS FOR USE THEREOF
;  FILE REFERENCE: DEKM:166
;  CURRENT APPLICATION NUMBER: US/09/850,964
;  CURRENT FILING DATE:  2001-05-07
;  PRIOR APPLICATION NUMBER: 09/312,038
;  PRIOR FILING DATE: 1999-05-14
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2289
;   TYPE: DNA
;   ORGANISM: Zea mays/rice
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (889)..(1825)
;   OTHER INFORMATION: N = A or C or G or T
US-09-850-964-3

  Query Match             32.5%;  Score 32.8;  DB 10;  Length 2289;
  Best Local Similarity   58.0%;  Pred. No. 1.3;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
             || |   |||  ||| ||| ||| || |||| ||  ||  | || |||| ||  | | |
Db     545 GCCGTAGCCGCCGCCACGGTTGCCACCGCCGTAGCCGCCACCACCACCACCGTAGCCACC 604

Qy      62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||   |   |   | | ||||  ||   | || | ||
Db     605 GCCGCCACCGTAGCCGCCGCCGCCACCGCCGTAGCCGCCG 644


RESULT 5
US-09-815-242-7987
; Sequence 7987, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
```

```
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7987
;   LENGTH: 1530
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1530)
US-09-815-242-7987

  Query Match             31.9%;  Score 32.2;  DB 10;  Length 1530;
  Best Local Similarity   66.7%;  Pred. No. 1.8;
  Matches   46;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           ||  |||||| ||| |||  | ||  ||       ||||||||   ||||| |||| |||
Db     952 GCCGCCGAGGGCGACGTGATCTCCGGCGGCAACTTCCACGCCGAACCGGTGGCGATGGCC 1011

Qy      65 GCGGTCAAC 73
           || | ||||
Db    1012 GCCGACAAC 1020


RESULT 6
US-09-815-242-7769
; Sequence 7769, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7769
;   LENGTH: 1395
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1395)
US-09-815-242-7769

  Query Match             31.7%;  Score 32;  DB 10;  Length 1395;
  Best Local Similarity   60.2%;  Pred. No. 2.1;
  Matches   53;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      14 GCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCGGTCAAC 73
           |||||| ||  ||||  |  |   | |||   |||| ||||| ||||| |||||||   |
Db     988 GCCGAGCTGATCGCCGGCAAGCGTCGCCAGTTCGCCCCGGTGGCGATCCCCGCGGTGTGC 1047

Qy      74 GGTCCTGACGCCGTCGTGATCGCTGGCG 101
              | ||  | |  ||| | |  | ||
Db    1048 TTCACCGATCCGGAAGTGGTGGTCGCCG 1075


RESULT 7
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-36

  Query Match             31.7%;  Score 32;  DB 10;  Length 4041;
  Best Local Similarity   79.2%;  Pred. No. 2;
  Matches   38;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      54 TGTCGATCGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||| |||||||| |||  ||||  ||||  |  | ||||
Db    2132 TGTCGATCGCCGCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCG 2179


RESULT 8
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
```

```
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             31.7%;  Score 32;  DB 10;  Length 36778;
  Best Local Similarity   79.2%;  Pred. No. 1.8;
  Matches   38;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy      54 TGTCGATCGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           |||||||||||||| |||||||| |||  ||||  |||| |   | ||||
Db   33913 TGTCGATCGCCGCCGTCAACGGGCCTACCGCCACCGTGGTTTCGGGCG 33960


RESULT 9
US-10-044-090-309
; Sequence 309, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 309
;   LENGTH: 2724
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 300437.18
;   NAME/KEY: unsure
;   LOCATION: 2060
;   OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-309

  Query Match             31.5%;  Score 31.8;  DB 12;  Length 2724;
  Best Local Similarity   67.2%;  Pred. No. 2.3;
  Matches   45;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       3 CAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCG 62
           | ||| |||||||||| |    |||||||| ||  |   |||||  ||| | | | ||
Db     601 CCGCGGCCGAGGCCGCGCCCCCCGCCTCCGCCGCCACTATCGCCGCGGCGGCGGCCACCG 660

Qy      63 CCGCGGT 69
           |||  ||
Db     661 CCGTCGT 667


RESULT 10
US-09-821-167-1/c
; Sequence 1, Application US/09821167
; Patent No. US20020015989A1
; GENERAL INFORMATION:
;  APPLICANT: Hosted Jr., Thomas J.
;  APPLICANT:  Horan, Ann C.
;  TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria
;  TITLE OF INVENTION:  plasmid pMR2 and Vectors Made Therefrom
;  FILE REFERENCE: IN01149Q
;  CURRENT APPLICATION NUMBER: US/09/821,167
;  CURRENT FILING DATE:  2001-03-29
;  PRIOR APPLICATION NUMBER: US 60/194,461
;  PRIOR FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 11188
;   TYPE: DNA
;   ORGANISM: Micromonospora rosaria
;   FEATURE:
;   NAME/KEY: RBS
;   LOCATION: (6055)..(6059)
;   NAME/KEY: RBS
;   LOCATION: (6391)..(6394)
;   NAME/KEY: RBS
;   LOCATION: (8084)..(8088)
;   NAME/KEY: RBS
;   LOCATION: (9834)..(9837)
;   NAME/KEY: RBS
;   LOCATION: (10010)..(10012)
US-09-821-167-1

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 11188;
  Best Local Similarity   59.3%;  Pred. No. 2.1;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      11 GAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCCGCGGTC 70
           |||||||||  ||  |   ||| | ||      || ||||||||||| | |      | |
Db    1507 GAGGCCGAGCGGGGTGAAGCCGCGCCAGAGGCAGCGGCCACGGTGTGGCTTTTGAAGCTG 1448

Qy      71 AACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           | || |  ||||| | ||| | |  || |||
Db    1447 AGCGCTGGTGACGACATCGGGCTATCTCGCG 1417


RESULT 11
US-09-880-107-2102/c
; Sequence 2102, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2102
;   LENGTH: 2038
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 J04029
US-09-880-107-2102

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 2038;
  Best Local Similarity   57.0%;  Pred. No. 3.2;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
           || ||  |||  ||||  |  |||||| |||| ||  ||  ||||||| | |   |   |
Db    1561 GCTGCTGCCGCCGCCGGAGCTGCCGCCCCCGTAGCCGCCGCCGCCGCCGCCGAAACTGCC 1502

Qy      62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || | ||   ||   |||   | | ||   | || | ||
Db    1501 ACCACCGTAGCCGCCGCTGGAACTGCCGCCGTGGCCGCCG 1462


RESULT 12
US-09-861-289-34/c
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
```

```
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 4689;
  Best Local Similarity   57.0%;  Pred. No. 3.1;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
          || ||   |||||  ||||||  || ||||||||    |||  | ||    |    |
Db   3444 GCGGCTGACGAGGAGGAGGTGTTCGGCTCCGTGGTGGGCCATCCAGCGTCGGCGTGGCT 3385

Qy     62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||  || |  |||| | | ||  |  | |  || || ||
Db   3384 GCCGAGGGCTCCGGTGCCGCCGGTGATGAGGACGGTGCCG 3345


RESULT 13
US-09-880-107-3699/c
; Sequence 3699, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3699
;   LENGTH: 6483
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 X14487
US-09-880-107-3699

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 6483;
  Best Local Similarity   57.0%;  Pred. No. 3.1;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
          || ||  |||  ||||  |  |||||| |||| ||  ||  |||||||  | |   |   |
Db   5407 GCTGCTGCCGCCGCCGGAGCTGCCGCCCCCGTAGCCGCCGCCGCCGCCGCCGGAACTGCC 5348

Qy     62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           || | ||   ||  |||   | | ||   | || | ||
Db   5347 ACCACCGTAGCCGCCGCTGGAACTGCCGCCGTGGCCGCCG 5308


RESULT 14
US-09-861-289-30/c
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 13842;
  Best Local Similarity   57.0%;  Pred. No. 2.9;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
          || ||   |||||  ||||||  || ||||||||    |||  | ||    |    |
Db   6720 GCGGCTGACGAGGAGGAGGTGTTCGGCTCCGTGGTGGGCCATCCAGCGTCGGCGTGGCT 6661

Qy     62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||  || |  |||| | | ||  |  | |  || || ||
Db   6660 GCCGAGGGCTCCGGTGCCGCCGGTGATGAGGACGGTGCCG 6621


RESULT 15
US-09-861-289-5/c
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 36778;
  Best Local Similarity   57.0%;  Pred. No. 2.8;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
          || ||   |||||  ||||||  || ||||||||    |||  | ||    |    |
Db  30434 GCGGCTGACGAGGAGGAGGTGTTCGGCTCCGTGGTGGGCCATCCAGCGTCGGCGTGGCT 30375

Qy     62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
          ||||  || |  |||| | | ||  |  | |  || || ||
Db  30374 GCCGAGGGCTCCGGTGCCGCCGGTGATGAGGACGGTGCCG 30335


Search completed: November  5, 2002, 23:01:40
Job time : 104.254 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_27200_27300
Perfect score: 101
Sequence: 1 cgcagcgtccgaggccgagg.....cgccgtcgtgatcgctggcg 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41.6	41.2	507	10	AV434749	AV434749 AV434749
2	41.6	41.2	512	10	AV434343	AV434343 AV434343
3	41.6	41.2	520	10	AV431953	AV431953 AV431953
4	41.6	41.2	528	10	AV431696	AV431696 AV431696
5	41.6	41.2	530	10	AV431496	AV431496 AV431496
c 6	41	40.6	495	12	BF630328	BF630328 HVSMEb000
c 7	40.2	39.8	403	13	BM097135	BM097135 EBpi07_SQ
c 8	40.2	39.8	434	12	BF264131	BF264131 HV_CEa000
c 9	39.4	39.0	883	10	BE215355	BE215355 HV_CEb000
c 10	39.2	38.8	528	13	BI778790	BI778790 EBro01_SQ
c 11	38.6	38.2	254	14	BQ759737	BQ759737 EBpi05_SQ
c 12	38.6	38.2	349	14	BQ764218	BQ764218 EBan01_SQ
c 13	38.6	38.2	375	14	BQ765195	BQ765195 EBro03_SQ
c 14	38.6	38.2	402	14	BQ766713	BQ766713 EBro08_SQ
c 15	38.6	38.2	462	13	BM370378	BM370378 EBro08_SQ
c 16	38.6	38.2	580	10	BE195010	BE195010 HVSMEh008
c 17	38.6	38.2	616	13	BI779974	BI779974 EBma08_SQ
18	38.6	38.2	620	12	BG300836	BG300836 HVSMEb001
c 19	38.6	38.2	767	12	BF630685	BF630685 HVSMEb001
c 20	38.6	38.2	833	13	BI952166	BI952166 HVSMEm000
c 21	38.6	38.2	907	12	BF264369	BF264369 HV_CEa000
c 22	38.4	38.0	488	10	BE237834	BE237834 894038E03
c 23	38.4	38.0	506	14	BQ814915	BQ814915 1030046F0
24	38.2	37.8	517	10	BE361647	BE361647 DG1_81_A0
c 25	38.2	37.8	547	12	BG300222	BG300222 HVSMEb001
c 26	37.8	37.4	212	14	BQ761467	BQ761467 EBro07_SQ
27	37.8	37.4	439	9	AU089840	AU089840 AU089840
c 28	37.8	37.4	635	14	BQ739816	BQ739816 HB03B07 H
c 29	37.8	37.4	667	10	BE559054	BE559054 HV_CEb002
c 30	37.8	37.4	705	10	BE559066	BE559066 HV_CEb002
c 31	37.8	37.4	807	12	BF264367	BF264367 HV_CEa000
c 32	37.8	37.4	812	12	BF065042	BF065042 HV_CEb002
c 33	37.6	37.2	251	13	BM369860	BM369860 EBem07_SQ
c 34	37.6	37.2	607	10	AW982877	AW982877 HVSMEg000
c 35	37.6	37.2	952	12	BG343478	BG343478 HVSMEg000
c 36	37.2	36.8	763	12	BF629013	BF629013 HVSMEb000
c 37	37	36.6	110	14	BQ766240	BQ766240 EBro08_SQ
c 38	37	36.6	377	13	BM097226	BM097226 EBpi07_SQ
c 39	37	36.6	392	14	BQ760747	BQ760747 EBro03_SQ
c 40	37	36.6	499	13	BM370579	BM370579 EBro08_SQ
c 41	37	36.6	542	10	BE489966	BE489966 WHE0363_H
c 42	37	36.6	640	14	BQ841536	BQ841536 WHE4212_B
c 43	37	36.6	718	12	BG367529	BG367529 HVSMEi001
c 44	37	36.6	871	12	BG345046	BG345046 HVSMEg001
c 45	37	36.6	909	14	BQ213316	BQ213316 AGENCOURT

ALIGNMENTS

RESULT 1
AV434749
LOCUS AV434749 507 bp mRNA linear EST 23-AUG-2000
DEFINITION AV434749 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone PM047f04_r 5', mRNA sequence.
ACCESSION AV434749
VERSION AV434749.1 GI:8589974
KEYWORDS EST.
SOURCE Porphyra yezoensis.
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
REFERENCE 1 (bases 1 to 507)
AUTHORS Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and Tabata,S.
TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1. .507
/organism="Porphyra yezoensis"
/strain="TU-1"
/db_xref="taxon:2788"

```
                     /clone="PM047f04_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       64 a    209 c    153 g     81 t
ORIGIN

  Query Match             41.2%;  Score 41.6;  DB 10;  Length 507;
  Best Local Similarity   67.0%;  Pred. No. 4.2;
  Matches   59;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           ||| ||| | | ||||  | ||| ||||  ||||||  ||  |||||||||| |   |
Db     229 CGCGGCGGCGGTGGCCATGCTGGTCGCCGTCGTGGCCGCCACCGCCGCCACCGGCGCCGC 288

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTC 88
           | |||| |||  || |||||  |||| |
Db     289 CACCGCAGTCGCCGCTCCTGCTGCCGCC 316


RESULT 2
AV434343
LOCUS       AV434343                 512 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV434343 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PM041h03_r 5', mRNA sequence.
ACCESSION   AV434343
VERSION     AV434343.1  GI:8589568
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Banglophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PM041h03_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       66 a    209 c    153 g     84 t
ORIGIN

  Query Match             41.2%;  Score 41.6;  DB 10;  Length 512;
  Best Local Similarity   67.0%;  Pred. No. 4.2;
  Matches   59;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           ||| ||| | | ||||  | ||| ||||  ||||||  ||  |||||||||| |   |
Db     231 CGCGGCGGCGGTGGCCATGCTGGTCGCCGTCGTGGCCGCCACCGCCGCCACCGGCGCCGC 290

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTC 88
           | |||| |||  || |||||  |||| |
Db     291 CACCGCAGTCGCCGCTCCTGCTGCCGCC 318


RESULT 3
AV431953
LOCUS       AV431953                 520 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV431953 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PM008a11_r 5', mRNA sequence.
ACCESSION   AV431953
VERSION     AV431953.1  GI:8587178
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 520)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .520
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PM008a11_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       66 a    213 c    157 g     84 t
ORIGIN

  Query Match             41.2%;  Score 41.6;  DB 10;  Length 520;
  Best Local Similarity   67.0%;  Pred. No. 4.2;
  Matches   59;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           ||| ||| | | ||||  | ||| ||||  ||||||  ||  |||||||||| |   |
Db     235 CGCGGCGGCGGTGGCCATGCTGGTCGCCGTCGTGGCCGCCACCGCCGCCACCGGCGCCGC 294

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTC 88
           | |||| |||  || |||||  |||| |
Db     295 CACCGCAGTCGCCGCTCCTGCTGCCGCC 322


RESULT 4
AV431696
LOCUS       AV431696                 528 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV431696 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PM004c09_r 5', mRNA sequence.
ACCESSION   AV431696
VERSION     AV431696.1  GI:8586921
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
```

```
                     /db_xref="taxon:2788"
                     /clone="PM004c09_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       71 a    211 c    160 g     86 t
ORIGIN

  Query Match             41.2%;  Score 41.6;  DB 10;  Length 528;
  Best Local Similarity   67.0%;  Pred. No. 4.2;
  Matches   59;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||| ||| | | ||||  | ||| ||||  ||||||  ||  |||||||||| |   |
Db    206 CGCGGCGGCGGTGGCCATGCTGGTCGCCGTCGTGGCCGCCACCGCCGCCACCGGCGCCGC 265

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTC 88
          | |||| |||  || |||||  |||| |
Db    266 CACCGCAGTCGCCGCTCCTGCTGCCGCC 293


RESULT 5
AV431496
LOCUS       AV431496                 530 bp    mRNA    linear   EST 23-AUG-2000
DEFINITION  AV431496 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
            PM001c08_r 5', mRNA sequence.
ACCESSION   AV431496
VERSION     AV431496.1  GI:8586721
KEYWORDS    EST.
SOURCE      Porphyra yezoensis.
  ORGANISM  Porphyra yezoensis
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and
            Tabata,S.
  TITLE     Generation of 10,154 expressed sequence tags from a leafy
            gametophyte of a marine red alga, Porphyra yezoensis
  JOURNAL   DNA Res. 7, 223-227 (2000)
  MEDLINE   20363100
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .530
                     /organism="Porphyra yezoensis"
                     /strain="TU-1"
                     /db_xref="taxon:2788"
                     /clone="PM001c08_r"
                     /clone_lib="Porphyra yezoensis TU-1"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       67 a    214 c    167 g     82 t
ORIGIN

  Query Match             41.2%;  Score 41.6;  DB 10;  Length 530;
  Best Local Similarity   67.0%;  Pred. No. 4.2;
  Matches   59;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
          ||| ||| | | ||||  | ||| ||||  ||||||  ||  |||||||||| |   |
Db    190 CGCGGCGGCGGTGGCCATGCTGGTCGCCGTCGTGGCCGCCACCGCCGCCACCGGCGCCGC 249

Qy     61 CGCCGCGGTCAACGGTCCTGACGCCGTC 88
          | |||| |||  || |||||  |||| |
Db    250 CACCGCAGTCGCCGCTCCTGCTGCCGCC 277


RESULT 6
BF630328/c
LOCUS       BF630328                 495 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEb0008M06f Hordeum vulgare seedling shoot EST library
            HVcDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone
            HVSMEb0008M06f, mRNA sequence.
ACCESSION   BF630328
VERSION     BF630328.2  GI:13090843
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 495)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex drought-stressed seedling shoot cDNA
            library
  JOURNAL   Unpublished (2001)
COMMENT     On Dec 19, 2000 this sequence version replaced gi:11894486.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 395
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 484.
FEATURES             Location/Qualifiers
     source          1. .495
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEb0008M06f"
                     /clone_lib="Hordeum vulgare seedling shoot EST library
                     HVcDNA0002 (Dehydration stress)"
                     /tissue_type="Seedling shoot"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedlings were
                     incubated at 90% RH for 24 hr. Shoots were then harvested,
                     total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, 600000 pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids.
                     These steps were performed in the TJ Close laboratory at
                     the University of California, Riverside (Choi, Close,
                     Fenton). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       78 a    146 c    198 g     73 t
ORIGIN

  Query Match             40.6%;  Score 41;  DB 12;  Length 495;
  Best Local Similarity   63.9%;  Pred. No. 5.7;
  Matches   62;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;
```

```
Qy       5 GCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATCGCC 64
           || |||  |||||  || ||| ||||| ||||  || | |||||||| | | ||   |||
Db     494 GCCTCCATGGCCGCCGTAGCCCCCTCCCTGGCCGCCGAAGCCGCCACCGCGCCGTAAGCC 435

Qy      65 GCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             | | || ||   |    |||| |||   ||| | ||
Db     434 TCCGCCACCGCGGCGCTGGCCGCCGTAGCCGCCGCCG 398


RESULT 7
BM097135/c
LOCUS       BM097135                 403 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBpi07_SQ002_I07_R pistil, 12 DPA, no treatment, cv Optic, EBpi07
            Hordeum vulgare cDNA clone EBpi07_SQ002_I07 5', mRNA sequence.
ACCESSION   BM097135
VERSION     BM097135.2  GI:21946243
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 403)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 21, 2001 this sequence version replaced gi:17027705.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .403
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBpi07_SQ002_I07"
                     /clone_lib="pistil, 12 DPA, no treatment, cv Optic,
                     EBpi07"
                     /tissue_type="pistil"
                     /dev_stage="12 DPA"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from pistils dissected from developing grains (12
                     days post anthesis) in glasshouse grown barley plants.
                     Developed as part of the barley transcriptome resources of
                     BBSRC/SEERAD funded cereal IGF (Investigating Gene
                     Function) project."
BASE COUNT       50 a    129 c    175 g     49 t
ORIGIN

  Query Match             39.8%;  Score 40.2;  DB 13;  Length 403;
  Best Local Similarity   62.4%;  Pred. No. 8.7;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||| |||||  || ||| ||||| ||||  ||   ||||||||  |   |  |
Db     206 CACCGCCTCCGTGGCCGCAGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCGCAGT 147

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             ||| | | || ||   |    |||| |||   ||| | ||
Db     146 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 106


RESULT 8
BF264131/c
```

```
LOCUS       BF264131                 434 bp    mRNA    linear   EST 23-OCT-2001
DEFINITION  HV_CEa0008J19f Hordeum vulgare seedling green leaf EST library
            HVcDNA0004 (Blumeria challenged) Hordeum vulgare cDNA clone
            HV_CEa0008J19f, mRNA sequence.
ACCESSION   BF264131
VERSION     BF264131.2  GI:13261289
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Wei,F., Begum,D.,
            Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi
            ,D.W., Fenton,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected incompatible (Mla13)
            seedling leaf cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Nov 17, 2000 this sequence version replaced gi:11195125.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 181
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 422.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Hordeum vulgare"
                     /cultivar="CI16155 (Mla13)"
                     /db_xref="taxon:4513"
                     /clone="HV_CEa0008J19f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVcDNA0004 (Blumeria challenged)"
                     /tissue_type="seedling green leaf"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
                     C.I. 16155 (Mla13) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were challenged with isolate A27 (AvrMla13
                     ) of Blumeria graminis f. sp. hordei, and leaves were
                     harvested 20 and 24 hr post-inoculation and snap frozen;
                     uninoculated leaves were harvested 20 hr post-inoculation
                     (Wei, Wise). In the TJ Close lab at the University of
                     California, Riverside, total RNA was prepared from each
                     sample pool, equal quantities of all three RNA pools were
                     combined, poly(A) RNA was purified from the mixture, one
                     cDNA library was made, and 1 million pfu were in vivo
                     excised to give pBluescript SK(-) cDNA phagemids (Choi,
                     Close). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       67 a    129 c    181 g     57 t
ORIGIN

  Query Match             39.8%;  Score 40.2;  DB 12;  Length 434;
```

```
  Best Local Similarity   62.4%;  Pred. No. 8.6;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  |||||  || ||| ||||| ||||  ||   |||||||| |  ||  |
Db     230 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCTCTGT 171

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
             ||| | | || ||   |    |||| |||   ||| | ||
Db     170 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 130


RESULT 9
BE215355/c
LOCUS       BE215355                883 bp    mRNA    linear   EST 23-OCT-2001
DEFINITION  HV_CEb0006J05f Hordeum vulgare seedling green leaf EST library
            HVcDNA0005 (Blumeria challenged) Hordeum vulgare cDNA clone
            HV_CEb0006J05f, mRNA sequence.
ACCESSION   BE215355
VERSION     BE215355.2  GI:13264386
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Wei,F., Begum,D.,
            Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi
            ,D.W., Fenton,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected incompatible (Mla6) seedling
            leaf cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 3, 2000 this sequence version replaced gi:8902967.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 312
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Hordeum vulgare"
                     /cultivar="CI16151 (Mla6)"
                     /db_xref="taxon:4513"
                     /clone="HV_CEb0006J05f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVcDNA0005 (Blumeria challenged)"
                     /tissue_type="seedling green leaf"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     C.I. 16151 (Mla6) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were challenged with isolate 5874 (AvrMla6
                     ) of Blumeria graminis f. sp. hordei, and leaves were
                     harvested 20 and 24 hr post-inoculation and snap frozen;
                     uninoculated leaves were harvested 20 hr post-inoculation
                     (Wei, Wise). In the TJ Close lab at the University of
                     California, Riverside, total RNA was prepared from each
                     sample pool, equal quantities of all three RNA pools were
                     combined, poly(A) RNA was purified from the mixture, one
                     primary unamplified cDNA library was made, and 1 million
                     pfu were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids (Choi, Close). Phagemids were plated and picked
                     at the Clemson University Genomics Institute (CUGI) (Begum
                     , Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      130 a    199 c    396 g    156 t      2 others
ORIGIN

  Query Match             39.0%;  Score 39.4;  DB 10;  Length 883;
  Best Local Similarity   62.9%;  Pred. No. 13;
  Matches   61;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 GCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGATC 61
           || ||  |||  ||||  | ||| ||||||||  ||  ||  |||||||  | ||  |   |
Db     354 GCCGCCGCCGGAGCCGCCGCGGCTGCCTCCGTAGCCGCCGCCGCCGCCGCCGTAGCCGCC 295

Qy      62 GCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTG 98
           |||||  ||||  | | | |  |||  ||    ||| |
Db     294 GCCGCCGTCACGGCTGCCGTAGCCACCGCCGCCGCCG 258


RESULT 10
BI778790/c
LOCUS       BI778790                528 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBro01_SQ001_I06_R root, 3 week, hydroponic grown, no treatment, cv
            Optic, EBro01 Hordeum vulgare cDNA clone EBro01_SQ001_I06 5', mRNA
            sequence.
ACCESSION   BI778790
VERSION     BI778790.2  GI:21946503
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Sep 26, 2001 this sequence version replaced gi:15781682.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro01_SQ001_I06"
                     /clone_lib="root, 3 week, hydroponic grown, no treatment,
                     cv Optic, EBro01"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old hydroponically grown
                     unstressed barley plants.  Developed as part of the barley
                     transcriptome resources of BBSRC/SEERAD funded cereal IGF
                     (Investigating Gene Function) project."
BASE COUNT       81 a    151 c    215 g     79 t      2 others
```

```
ORIGIN

  Query Match             38.8%;  Score 39.2;  DB 13;  Length 528;
  Best Local Similarity   61.4%;  Pred. No. 14;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCC TCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  |||||  || ||| ||||| ||||  ||   |||||||| |   | |
Db     501 CACCGCCTCCCTGGCCGCCGTAGCCACC TCCCTGGCCGCCGTAGCCGCCACCGCCGCNGT 442

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTC GTGATCGCTGGCG 101
           ||| | | || ||   |    |||| | |   ||| | ||
Db     441 AGCCTCCGCCACCGCCGCGCTGGCCGCC GTAGCCGCCGCCG 401


RESULT 11
BQ759737/c
LOCUS       BQ759737                 254 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBpi05_SQ004_B08_R pistil, 8 DPA, no treatment, cv Optic, EBpi05
            Hordeum vulgare cDNA clone EBpi05_SQ004_B08 5', mRNA sequence.
ACCESSION   BQ759737
VERSION     BQ759737.1  GI:21968209
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 254)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .254
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBpi05_SQ004_B08"
                     /clone_lib="pistil, 8 DPA, no treatment, cv Optic, EBpi05"
                     /tissue_type="pistil"
                     /dev_stage="8 DPA"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from pistils dissected from developing grains (8
                     days post anthesis) in glasshouse grown barley plants.
                     Developed as part of the barley transcriptome resources of
                     BBSRC/SEERAD funded cereal IGF (Investigating Gene
                     Function) project."
BASE COUNT       23 a     74 c    133 g     24 t
ORIGIN

  Query Match             38.2%;  Score 38.6;  DB 14;  Length 254;
  Best Local Similarity   61.4%;  Pred. No. 20;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  |||||  || ||| ||||| ||||  ||   |||||||| |   | |
Db     105 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCGCCGT 46

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           ||| | | || ||   |    |||| |||   ||| | ||
Db      45 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 5


RESULT 12
BQ764218/c
LOCUS       BQ764218                 349 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBan01_SQ005_O06 _R anther, yellow stage, no treatment, cv Optic,
            EBan01 Hordeum vulgare cDNA clone EBan01_SQ005_O06 5', mRNA
            sequence.
ACCESSION   BQ764218
VERSION     BQ764218.1  GI:21972690
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBan01_SQ005_O06"
                     /clone_lib="anther, yellow stage, no treatment, cv Optic,
                     EBan01"
                     /tissue_type="anther"
                     /dev_stage="yellow stage"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from anthers dissected from developing flowers in
                     glasshouse grown barley plants. Developed as part of the
                     barley transcriptome resources of BBSRC/SEERAD funded
                     cereal IGF (Investigating Gene Function) project."
BASE COUNT       46 a     92 c    158 g     53 t
ORIGIN

  Query Match             38.2%;  Score 38.6;  DB 14;  Length 349;
  Best Local Similarity   61.4%;  Pred. No. 20;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  |||||  || ||| ||||| ||||  ||   |||||||| |   | |
Db     325 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCGCCGT 266

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
           ||| | | || ||   |    |||| |||   ||| | ||
Db     265 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 225


RESULT 13
BQ765195/c
LOCUS       BQ765195                 375 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBro03_SQ006_A18_R root, 3 week, waterlogged, cv Optic, EBro03
            Hordeum vulgare cDNA clone EBro03_SQ006_A18 5', mRNA sequence.
ACCESSION   BQ765195
VERSION     BQ765195.1  GI:21973669
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 375)
```

```
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro03_SQ006_A18"
                     /clone_lib="root, 3 week, waterlogged, cv Optic, EBro03"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old waterlogged barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT       58 a    117 c    161 g     39 t
ORIGIN

  Query Match             38.2%;  Score 38.6;  DB 14;  Length 375;
  Best Local Similarity   61.4%;  Pred. No. 20;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  ||||  || ||| ||||| ||||  ||   |||||||| |   | |
Db     350 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCCCCGT 291

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
            ||| | | || ||   |    |||| |||   ||| | ||
Db     290 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 250


RESULT 14
BQ766713/c
LOCUS       BQ766713                402 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBro08_SQ006_M20 _R root, 3 week, drought-stressed, cv Optic,
            EBro08 Hordeum vulgare cDNA clone EBro08_SQ006_M20 5', mRNA
            sequence.
ACCESSION   BQ766713
VERSION     BQ766713.1  GI:21975187
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 402)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .402
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro08_SQ006_M20"
                     /clone_lib="root, 3 week, drought-stressed, cv Optic,
                     EBro08"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old drought stressed barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT       60 a    127 c    169 g     46 t
ORIGIN

  Query Match             38.2%;  Score 38.6;  DB 14;  Length 402;
  Best Local Similarity   61.4%;  Pred. No. 20;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || |||  ||||  || ||| ||||| ||||  ||   |||||||| |   | |
Db     375 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCGCCGT 316

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCGTGATCGCTGGCG 101
            ||| | | || ||   |    |||| |||   ||| | ||
Db     315 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 275


RESULT 15
BM370378/c
LOCUS       BM370378                462 bp    mRNA    linear   EST 23-JUL-2002
DEFINITION  EBro08_SQ004_A15_R root, 3 week, drought-stressed, cv Optic, EBro08
            Hordeum vulgare cDNA clone EBro08_SQ004_A15 5', mRNA sequence.
ACCESSION   BM370378
VERSION     BM370378.2  GI:21951484
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     On Jan 10, 2002 this sequence version replaced gi:18113768.
            Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk
            All sequence has a Phred quality score of 20 or over
            Seq primer: M13 reverse.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro08_SQ004_A15"
                     /clone_lib="root, 3 week, drought-stressed, cv Optic,
                     EBro08"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old drought stressed barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
```

Gene Function) project."
BASE COUNT 53 a 136 c 185 g 88 t
ORIGIN

Query Match 38.2%; Score 38.6; DB 13; Length 462;
Best Local Similarity 61.4%; Pred. No. 19;
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```
Qy       1 CGCAGCGTCCGAGGCCGAGGTGGCCGCCTCCGTGGCACCCCACGCCGCCACGGTGTCGAT 60
           | | || ||| |||||  || ||| || || ||||  ||   |||||||| |   | |
Db     123 CACCGCCTCCCTGGCCGCCGTAGCCACCTCCCTGGCCGCCGTAGCCGCCACCGCCGCCGT 64

Qy      61 CGCCGCGGTCAACGGTCCTGACGCCGTCCTGATCGCTGGCG 101
           ||| | | || ||   |    |||| |||   ||| | ||
Db      63 AGCCTCCGCCACCGCCGCGCTGGCCGCCGTAGCCGCCGCCG 23
```

Search completed: November 5, 2002, 16:18:13
Job time : 1117.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_24950_25050
Perfect score: 101
Sequence: 1 cgtcgctcgggtgctgggct.....gatggcggtggagatccgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	41.8	41.4	11220	10	US-09-861-289-32	Sequence 32, Appl
2	41.8	41.4	15872	10	US-09-861-289-1	Sequence 1, Appli
3	41.8	41.4	36778	10	US-09-861-289-5	Sequence 5, Appli
4	35.4	35.0	4689	10	US-09-861-289-34	Sequence 34, Appl
5	34.4	34.1	3209	10	US-09-925-301-474	Sequence 474, App
6	31.2	30.9	435	10	US-09-738-973-405	Sequence 405, App
7	31.2	30.9	518	10	US-09-833-790-317	Sequence 317, App
8	31.2	30.9	13842	10	US-09-861-289-30	Sequence 30, Appl
9	31.2	30.9	110096	10	US-09-880-107-1542	Sequence 1542, Ap
10	29.6	29.3	1958	10	US-09-791-406-3	Sequence 3, Appli
11	29	28.7	3624	10	US-09-927-112-3	Sequence 3, Appli
12	29	28.7	4635	10	US-09-927-112-1	Sequence 1, Appli
13	28.8	28.5	2073	10	US-09-815-242-7742	Sequence 7742, Ap
c 14	28.6	28.3	14800	10	US-09-954-456-1601	Sequence 1601, Ap
c 15	28.2	27.9	4689	10	US-09-861-289-34	Sequence 34, Appl
c 16	28.2	27.9	36778	10	US-09-861-289-5	Sequence 5, Appli
17	28	27.7	385	10	US-09-960-352-1052	Sequence 1052, Ap
c 18	27.8	27.5	1938	10	US-09-815-242-7950	Sequence 7950, Ap
19	27.2	26.9	1533	10	US-09-815-242-7755	Sequence 7755, Ap
c 20	27.2	26.9	2080	10	US-09-951-768A-1	Sequence 1, Appli
21	27	26.7	269	10	US-09-867-701-3630	Sequence 3630, Ap
22	27	26.7	271	10	US-09-867-701-5599	Sequence 5599, Ap
23	27	26.7	312	10	US-09-867-701-3198	Sequence 3198, Ap
24	27	26.7	1832	10	US-09-925-302-58	Sequence 58, Appl
25	26.8	26.5	4446	10	US-09-815-242-7983	Sequence 7983, Ap
26	26.6	26.3	235	10	US-09-923-876-2937	Sequence 2937, Ap
27	26.4	26.1	434	10	US-09-960-352-6062	Sequence 6062, Ap
28	26.4	26.1	436	10	US-09-960-352-13405	Sequence 13405, A
c 29	26.4	26.1	1969	10	US-09-864-761-4449	Sequence 4449, Ap
c 30	26.4	26.1	2004	10	US-09-864-761-21197	Sequence 21197, A
31	25.8	25.5	4270	10	US-09-880-107-2427	Sequence 2427, Ap
32	25.4	25.1	1281	10	US-09-861-289-19	Sequence 19, Appl
33	25.4	25.1	2001	10	US-09-804-551B-15	Sequence 15, Appl
34	25.4	25.1	3145	10	US-09-764-869-2190	Sequence 2190, Ap
35	25.4	25.1	13613	10	US-09-861-289-3	Sequence 3, Appli
c 36	25.2	25.0	259	10	US-09-923-876-3088	Sequence 3088, Ap
c 37	25.2	25.0	377	10	US-09-864-761-21329	Sequence 21329, A
38	25.2	25.0	762	10	US-09-815-242-7745	Sequence 7745, Ap
c 39	25.2	25.0	1647	10	US-09-764-870-144	Sequence 144, App
40	25	24.8	470	10	US-09-864-761-5495	Sequence 5495, Ap
c 41	25	24.8	733	10	US-09-833-381-69	Sequence 69, Appl
42	25	24.8	1198	10	US-09-791-171-15	Sequence 15, Appl
43	25	24.8	1704	10	US-09-815-242-7788	Sequence 7788, Ap
44	25	24.8	1779	10	US-09-903-383-1	Sequence 1, Appli
45	25	24.8	1785	10	US-09-732-357A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Query Match 41.4%; Score 41.8; DB 10; Length 11220;
Best Local Similarity 63.4%; Pred. No. 0.00045;
Matches 64; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db    4362 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 4421

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db    4422 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 4462
```

RESULT 2
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/851,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             41.4%;  Score 41.8;  DB 10;  Length 15872;
  Best Local Similarity   63.4%;  Pred. No. 0.00047;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||    || || |  | |  |||||||| |    |||| |  ||  |    |||||
Db    2833 CGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAA 2892

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||| |||| |||||||  | ||| | || ||| | || |
Db    2893 GGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGA 2933


RESULT 3
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             41.4%;  Score 41.8;  DB 10;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 0.0005;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db   20049 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 20108

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| |||||||
Db   20109 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 20149


RESULT 4
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             35.0%;  Score 35.4;  DB 10;  Length 4689;
  Best Local Similarity   59.4%;  Pred. No. 0.033;
  Matches   60;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||  ||     || || |||      |  |||    |  | | |  || ||     |  ||||
Db    4236 CGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCAC 4295

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
             ||||  ||| |||||||| ||| ||| ||| || ||||||
Db    4296 CGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGCTCCGCA 4336


RESULT 5
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
;   LENGTH: 3209
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (427)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474

  Query Match             34.1%;  Score 34.4;  DB 10;  Length 3209;
  Best Local Similarity   59.0%;  Pred. No. 0.064;
  Matches   59;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||| || |   | |||||| ||| ||||  |   ||  || | |  |  |      |
Db    1225 CGTGGCACACATYCTGGGCATCCGCGACTTGGCTGCTGTCAACCTGGACAGCTCACTGGC 1284

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
             || | |||||| |||||| | |||   |||||| | |||
Db    1285 GGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGCGC 1324


RESULT 6
US-09-738-973-405
; Sequence 405, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
```

```
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Fling, Steven P.
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Indirias, Carol Yoseph
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Elliot, Mark
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Kalos, Michael D.
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  THE THERAPY AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.475C9
;  CURRENT APPLICATION NUMBER: US/09/738,973
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS:  587
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 405
;   LENGTH: 435
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-738-973-405

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 435;
  Best Local Similarity   63.2%;  Pred. No. 0.48;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          |||  | | | |||| ||| |||||| || |  ||    ||| ||||   |    |||||
Db     75 CGTGCCGCTGCTGCTCGGCCTCCTCGGCCTGGCCGTCGCCGAGCCTGCCGTCTACTTCAA 134

Qy     61 TGAGCAGGGCCTCGAC 76
          ||||||   || |||
Db    135 GGAGCAGTTTCTGGAC 150


RESULT 7
US-09-833-790-317
; Sequence 317, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
;  APPLICANT: Lodes, Michael J.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Mohamath, Raodoh
;  APPLICANT:  Indirias, Carol Y.
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.512
;  CURRENT APPLICATION NUMBER: US/09/833,790
;  CURRENT FILING DATE:  2001-04-11
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 317
;   LENGTH: 518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-833-790-317

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 518;
  Best Local Similarity   63.2%;  Pred. No. 0.49;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          |||  | | | |||| ||| |||||| || |  ||    ||| ||||   |    |||||
Db     81 CGTGCCGCTGCTGCTCGGCCTCCTCGGCCTGGCCGTCGCCGAGCCTGCCGTCTACTTCAA 140

Qy     61 TGAGCAGGGCCTCGAC 76
          ||||||   || |||
Db    141 GGAGCAGTTTCTGGAC 156


RESULT 8
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 13842;
  Best Local Similarity   57.0%;  Pred. No. 0.65;
  Matches   57;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy      2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
          |||||    || || |||  |  |  |      |  | ||||  | |    ||| |||
Db  13249 GTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCGACGCCGAACGGGGCTTCCTC 13308

Qy     62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          || |  ||| |||||||  |||  || || ||  ||||||
Db  13309 GACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 13348


RESULT 9
US-09-880-107-1542
; Sequence 1542, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1542
;   LENGTH: 110096
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 AD000092
US-09-880-107-1542

  Query Match             30.9%;  Score 31.2;  DB 10;  Length 110096;
  Best Local Similarity   63.2%;  Pred. No. 0.78;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          |||  | | | |||| ||| |||||| || |  ||    ||| ||||   |    |||||
Db  85260 CGTGCCGCTGCTGCTCGGCCTCCTCGGCCTGGCCGTCGCCGAGCCTGCCGTCTACTTCAA 85319
```

```
Qy      61 TGAGCAGGGCCTCGAC 76
           ||||||   || |||
Db   85320 GGAGCAGTTTCTGGAC 85335


RESULT 10
US-09-791-406-3
; Sequence 3, Application US/09791406
; Patent No. US20020147165A1
; GENERAL INFORMATION:
;  APPLICANT:  C. Frank Bennett
;  APPLICANT:  Robert Rothlein
;  APPLICANT:  Takashi Kei Kishimoto
;  APPLICANT:  Lex M. Cowsert
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF CALRETICULIN EXPRESSION
;  FILE REFERENCE:  RTS-0097
;  CURRENT APPLICATION NUMBER: US/09/791,406
;  CURRENT FILING DATE:  2001-02-22
;  NUMBER OF SEQ ID NOS:  89
; SEQ ID NO 3
;   LENGTH: 1958
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (109)..(1362)
US-09-791-406-3

  Query Match             29.3%;  Score 29.6;  DB 10;  Length 1958;
  Best Local Similarity   61.8%;  Pred. No. 1.6;
  Matches   47;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |||  |   | |||| ||| |||||| || |  ||    ||| ||||   |    |||||
Db     120 CGTGCCGTTGCTGCTCGGCCTCCTCGGCCTGGCCGTCGCCGAGCCTGCCGTCTACTTCAA 179

Qy      61 TGAGCAGGGCCTCGAC 76
           ||||||   || |||
Db     180 GGAGCAGTTTCTGGAC 195


RESULT 11
US-09-927-112-3
; Sequence 3, Application US/09927112
; Patent No. US20020106774A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Meyers, Rachel
;  APPLICANT:  Silos-Santiago, Inmaculada
;  TITLE OF INVENTION: 32544, a novel human phospholipase C and
;  TITLE OF INVENTION:  uses thereof
;  FILE REFERENCE: 38155-20048.00
;  CURRENT APPLICATION NUMBER: US/09/927,112
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: US 60/246,808
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3624
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-927-112-3

  Query Match             28.7%;  Score 29;  DB 10;  Length 3624;
  Best Local Similarity   63.8%;  Pred. No. 2.6;
  Matches   44;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      29 CCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAGCAGGGCCTCGACTCGTTGATGGCG 88
           || |   ||| |||  |   | |||||  |  || | ||  | ||||| | ||  ||| |
Db    2299 CCCAAGCCGCGCGACTCCATGCTGGGGGACCGTGGGGAGATCATCGACCCCTTTGTGGAG 2358

Qy      89 GTGGAGATC 97
           |||||||||
Db    2359 GTGGAGATC 2367


RESULT 12
US-09-927-112-1
; Sequence 1, Application US/09927112
; Patent No. US20020106774A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Meyers, Rachel
;  APPLICANT:  Silos-Santiago, Inmaculada
;  TITLE OF INVENTION: 32544, a novel human phospholipase C and
;  TITLE OF INVENTION:  uses thereof
;  FILE REFERENCE: 38155-20048.00
;  CURRENT APPLICATION NUMBER: US/09/927,112
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: US 60/246,808
;  PRIOR FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 4635
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (435)...(4058)
US-09-927-112-1

  Query Match             28.7%;  Score 29;  DB 10;  Length 4635;
  Best Local Similarity   63.8%;  Pred. No. 2.7;
  Matches   44;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      29 CCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAGCAGGGCCTCGACTCGTTGATGGCG 88
           || |   ||| |||  |   | |||||  |  || | ||  | ||||| | ||  ||| |
Db    2733 CCCAAGCCGCGCGACTCCATGCTGGGGGACCGTGGGGAGATCATCGACCCCTTTGTGGAG 2792

Qy      89 GTGGAGATC 97
           |||||||||
Db    2793 GTGGAGATC 2801


RESULT 13
US-09-815-242-7742
; Sequence 7742, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
```

```
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7742
;   LENGTH: 2073
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(2073)
US-09-815-242-7742

  Query Match             28.5%;  Score 28.8;  DB 10;  Length 2073;
  Best Local Similarity   62.5%;  Pred. No. 2.9;
  Matches   45;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      29 CCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAGCAGGGCCTCGACTCGTTGATGGCG 88
           | ||||||| ||    | |  || | |||||| | || |  | ||| | |||||||  |
Db    1864 CAGAGCGCGATCCTGTCGGCCATCGTGTTCAACGCGCTGATCATCGTCGCGTTGATCCCA 1923

Qy      89 GTGGAGATCCGC 100
           |||   | |||
Db    1924 CTGGCCTTGCGC 1935


RESULT 14
US-09-954-456-1601/c
; Sequence 1601, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1601
;   LENGTH: 14800
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-1601

  Query Match             28.3%;  Score 28.6;  DB 10;  Length 14800;
  Best Local Similarity   55.6%;  Pred. No. 3.9;
  Matches   55;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           | ||   | | ||||| | || ||   ||| || |||||||||| |||   ||  |||
Db    1887 GGCGGGGGAGAGCTGGCCCTCGTCACTCCGTGCCCGCTCGATCTTGGCCCGGAATTCTTC 1828

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGC 100
           ||         |  |  | | | |||| |   || | |||
Db    1827 GATGGACTGGTGCAGGCCTCGGTGGCTGCCCAGCTGCGC 1789


RESULT 15
US-09-861-289-34/c
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             27.9%;  Score 28.2;  DB 10;  Length 4689;
  Best Local Similarity   59.3%;  Pred. No. 4.6;
  Matches   48;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           || ||  ||  ||| |||| |||  ||| |   || ||   |  |||||| |||   |
Db    4105 CGGCGAGCGCCTGCCGGGCCTCCGGGACGCCGTCGAGCAGAAGGCTGGGACGGGACACCG 4046

Qy      61 TGAGCAGGGCCTCGACTCGTT 81
           ||| |  || | |||  || |
Db    4045 TGAACGCGGGCGCGAACCGCT 4025


Search completed: November  5, 2002, 23:01:01
Job time : 126.254 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_24950_25050
Perfect score: 101
Sequence: 1 cgtcgctcgggtgctgggct.....gatggcggtggagatccgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	48.8	48.3	43280	2	US-08-804-227C-1	Sequence 1, Appli
	10	47.2	46.7	33529	4	US-09-144-085-3	Sequence 3, Appli
	11	41.8	41.4	11220	4	US-09-105-537-32	Sequence 32, Appl
	12	41.8	41.4	15872	4	US-09-105-537-1	Sequence 1, Appli
	13	41.8	41.4	36778	4	US-09-105-537-5	Sequence 5, Appli
	14	41.8	41.4	38506	3	US-09-320-878-19	Sequence 19, Appl
	15	41.4	41.0	80161	3	US-09-036-987A-1	Sequence 1, Appli
	16	41.4	41.0	80161	4	US-09-370-700-1	Sequence 1, Appli
c	17	40.2	39.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	18	38.8	38.4	11219	1	US-07-642-734C-1	Sequence 1, Appli
	19	38.8	38.4	11219	3	US-08-439-009A-1	Sequence 1, Appli
	20	38.6	38.2	20235	1	US-07-642-734C-3	Sequence 3, Appli
	21	38.6	38.2	20235	3	US-08-439-009A-3	Sequence 3, Appli
	22	37.4	37.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
	23	37.4	37.0	28958	1	US-08-456-837-6	Sequence 6, Appli
	24	37.4	37.0	28958	1	US-08-457-342-6	Sequence 6, Appli
	25	37.4	37.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
	26	37.4	37.0	28958	1	US-08-458-076A-6	Sequence 6, Appli
	27	37.4	37.0	28958	1	US-08-764-233A-4	Sequence 4, Appli
	28	37.4	37.0	28958	1	US-08-457-335A-6	Sequence 6, Appli
	29	37.4	37.0	28958	1	US-08-729-214-6	Sequence 6, Appli
	30	37.4	37.0	28958	3	US-09-028-934-6	Sequence 6, Appli
	31	37.4	37.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
	32	35.4	35.0	1681	4	US-09-434-288-7	Sequence 7, Appli
	33	35.4	35.0	4689	4	US-09-105-537-34	Sequence 34, Appl
	34	34.4	34.1	8460	1	US-08-469-005A-9	Sequence 9, Appli
	35	34.4	34.1	8519	4	US-09-261-907-1	Sequence 1, Appli
	36	34.4	34.1	44377	2	US-08-804-227C-7	Sequence 7, Appli
	37	34.4	34.1	44377	2	US-08-804-198-1	Sequence 1, Appli
	38	34.4	34.1	50937	4	US-09-428-517-1	Sequence 1, Appli
	39	34.4	34.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	40	34.4	34.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	41	34	33.7	444	4	US-09-060-756-658	Sequence 658, App
c	42	32.8	32.5	367	4	US-09-060-756-627	Sequence 627, App
	43	32.8	32.5	13987	2	US-08-804-227C-13	Sequence 13, Appl
	44	31.2	30.9	13842	4	US-09-105-537-30	Sequence 30, Appl
	45	30	29.7	23673	4	US-09-773-816-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 2.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 68750
;    TYPE: DNA
;    ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 68750
;    TYPE: DNA
;    ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;    LENGTH: 71989
;    TYPE: DNA
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24950 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 25009

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25010 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 25050


RESULT 9
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
```

```
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  317-276-2459
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  43280 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  816..14234
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  14351..19945
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  20010..31199
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31232..36067
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             48.3%;  Score 48.8;  DB 2;  Length 43280;
  Best Local Similarity   68.0%;  Pred. No. 6.3e-06;
  Matches   68;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           ||  |  | ||||||||||| |   | ||||  |||| | || || |||     ||||||
Db   13779 GTGTCGGGCGTGCTGGGCTACTCGGGCCCGACGCGGTGGAGCCCGGGCGCCCGTTCAAG 13838

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTCGAGATCCGCA 101
           || |  ||| |||||||| |||  || ||||||| | ||||
Db   13839 GATCTCGGCTTCGACTCGCTGACCGCCGTCGAGCTGCGCA 13878


RESULT 10
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144 085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             46.7%;  Score 47.2;  DB 4;  Length 33529;
  Best Local Similarity   69.6%;  Pred. No. 1.8e-05;
  Matches   64;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAG 64
           ||   |||||||||| |    || ||||||| ||| || || | |   || ||    ||
Db    9459 GCGGCGGTGCTGGGCGTGAAGGATCCGAGCGGGCTGGACCCGGAGCGAGGCTTTCTGGAC 9518

Qy      65 CAGGGCCTCGACTCGTTGATGGCGGTGGAGAT 96
```

```
           |  || || ||||||||||||||||||||| |
Db    9519 CTCGGGCTGGACTCGTTGATGGCGGTGGAGCT 9550


RESULT 11
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             41.4%;  Score 41.8;  DB 4;  Length 11220;
  Best Local Similarity   63.4%;  Pred. No. 0.00062;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db    4362 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 4421

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db    4422 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 4462


RESULT 12
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             41.4%;  Score 41.8;  DB 4;  Length 15872;
  Best Local Similarity   63.4%;  Pred. No. 0.00065;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||    || || |  | |  |||||||| |    |||| |  ||  |    |||||
Db    2833 CGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAA 2892

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
            |||| |||| |||||||  | ||| | || ||| | || |
Db    2893 GGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGA 2933


RESULT 13
```

```
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             41.4%;  Score 41.8;  DB 4;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 0.00073;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db   20049 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 20108

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db   20109 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 20149


RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             41.4%;  Score 41.8;  DB 3;  Length 38506;
  Best Local Similarity   63.4%;  Pred. No. 0.00073;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||

Db   18191 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 18250

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db   18251 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 18291


RESULT 15
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             41.0%;  Score 41.4;  DB 3;  Length 80161;
  Best Local Similarity   65.9%;  Pred. No. 0.0011;
  Matches   60;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      11 GTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAGCAGGGC 70
           ||||||||   |  |||  | |  || || |  ||   |  |  |||| | || |  ||
Db   44500 GTGCTGGGACGCGACGATGCCACGGCCCTGGCGCCATCGCGGCCGTTCCAGGAACTCGGA 44559

Qy      71 CTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||  ||||||||||||||| | ||||
Db   44560 TTCGACTCCTTGATGGCGGTGGAGCTGCGCA 44590


Search completed: November  5, 2002, 15:12:27
Job time : 189.737 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_24950_25050
Perfect score: 101
Sequence: 1 cgtcgctcgggtgctgggct.....gatggcggtggagatccgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
2	101	100.0	71989	21	AAA29349	Sorangium cellulos
3	48.8	48.3	43280	18	AAT80413	Tylactone synthase
4	47.2	46.7	33529	23	AAS17367	DNA sequence of S.
5	45.6	45.1	65140	22	AAD17184	Streptomyces nours
6	45.6	45.1	125401	22	AAD17186	Streptomyces nours
7	43	42.6	20394	22	AAF24892	Pimaricin biosynth
8	42.4	42.0	5676	19	AAV21186	Amycolatopsis medi
9	42.4	42.0	53789	19	AAV21187	Amycolatopsis medi
10	41.8	41.4	11220	21	AAZ87298	S. venezuelae macr
11	41.8	41.4	15872	18	AAT68715	Streptomyces venez
12	41.8	41.4	15872	21	AAZ87283	S. venezuelae vep
13	41.8	41.4	36778	21	AAZ87318	S. venezuelae pik
14	41.8	41.4	37948	21	AAZ87285	S. venezuelae pik
15	41.8	41.4	38506	21	AAA75633	Nucleotide sequenc
16	41.8	41.4	38506	21	AAZ56001	Recombinant cosmid
17	41.4	41.0	990	21	AAC55784	Type I polyketide
18	41.4	41.0	9513	22	AAF88337	S. spinosa DNA fra
19	41.4	41.0	12249	21	AAC55840	Complete Mitomycin
20	41.4	41.0	18331	21	AAC55857	Complete nucleotid
21	41.4	41.0	50000	22	AAF88313	S. spinosa DNA fra
22	41.4	41.0	50000	22	AAF88316	S. spinosa DNA fra
23	41.4	41.0	80161	20	AAZ21501	DNA fragment of Sa
c 24	40.2	39.8	77536	21	AAA14651	Nucleotide sequenc
25	40.2	39.8	109519	22	AAS08693	Micromonospora DNA
c 26	40.2	39.8	4403765	22	AAI99683	Mycobacterium tube
27	40	39.6	7788	22	AAF88335	S. spinosa DNA fra
28	40	39.6	50000	22	AAF88312	S. spinosa DNA fra
29	39.2	38.8	30690	21	AAA92301	S. avermitilis ave
30	39.2	38.8	30690	22	AAH79277	Streptomyces averm
31	38.8	38.4	29879	14	AAQ46806	eryA region of S.
32	38.2	37.8	77536	21	AAA14651	Nucleotide sequenc
33	37.6	37.2	47981	22	AAF30757	Micromonospora meg
34	37.4	37.0	28598	17	AAT06769	Sorangium cellulos
35	37.4	37.0	28958	18	AAT89956	Sorangium cellulos
36	37.4	37.0	28958	21	AAA75299	DNA sequence of So
37	37.4	37.0	49377	19	AAV05287	The soraphen biosy
38	36.6	36.2	3978	21	AAC55785	Type I polyketide
39	36	35.6	5088	22	AAF90038	Nucleotide sequenc
40	36	35.6	6459	22	AAF88336	S. spinosa DNA fra
41	36	35.6	34071	22	AAF90033	Nucleotide sequenc
c 42	36	35.6	42717	22	AAF90032	Nucleotide sequenc
43	35.4	35.0	1681	24	AAS18438	Contig 115 DNA enc
44	35.4	35.0	4689	21	AAZ87299	S. venezuelae macr
45	35	34.7	16767	22	AAF88339	S. spinosa DNA fra

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes EH6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24950 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 25009

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25010 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 25050


RESULT 3
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             48.3%;  Score 48.8;  DB 18;  Length 43280;
  Best Local Similarity   68.0%;  Pred. No. 4.1e-05;
  Matches   68;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           ||  |  | ||||||||||| |   | ||||  |||| | || || |||     ||||||
Db   13779 GTGTCGGGCGTGCTGGGCTACTCGGGCCCGGACGCGGTGGAGCCCGGGCGCCCGTTCAAG 13838

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           || |  ||| |||||||| |||  || ||||||  | ||||
Db   13839 GATCTCGGCTTCGACTCGCTGACCGCCGTGGAGCTGCGCA 13878


RESULT 4
AAS17367
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   12-MAR-2002  (first entry)
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
```

```
XX
OS   Sorangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT   -
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the DNA
CC   sequence of cosmid pKOS28-26 which encodes one or more domains of
CC   S. cellulosum PKS.
CC   Note: The present sequence is said to encode the functional domains
CC   of S. cellulosum PKS which correspond to domains or domain subsets of
CC   the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

  Query Match             46.7%;  Score 47.2;  DB 23;  Length 33529;
  Best Local Similarity   69.6%;  Pred. No. 0.00012;
  Matches   64;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAG 64
           ||   |||||||||| |    || ||||||| ||| || || | |   || ||    ||
Db    9459 GCGGCGGTGCTGGGCGTGAAGGATCCGAGCGGGCTGGACCCGGAGCGAGGCTTTCTGGAC 9518

Qy      65 CAGGGCCTCGACTCGTTGATGGCGGTGGAGAT 96
           |  || || ||||||||||||||||||||| |
Db    9519 CTCGGGCTGGACTCGTTGATGGCGGTGGAGCT 9550


RESULT 5
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
```

```
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 22;  Length 65140;
  Best Local Similarity   66.0%;  Pred  No. 0.00038;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           | |||   |||||| ||   |  ||||      ||| ||| | | |||     |||| |
Db   49388 GCCGCGGTGGTGCTCGGGCACGCCGACGCCTCGGCGATCGGTGCGGGGCGACAGTTCCAG 49447

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||  |||| |||||||| |||  ||||||||| ||||||
Db   49448 GAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCCGCA 49487


RESULT 6
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 22;  Length 125401;
  Best Local Similarity   66.0%;  Pred. No. 0.00041;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           | |||   |||||| ||   |  ||||      ||| ||| | | |||     |||| |
Db 109649 GCCGCGGTGGTGCTCGGGCACGCCGACGGCTCGGCGATCGGTGCGGGGCGACAGTTCCAG 109708

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||  |||| |||||||| |||  ||||||||| ||||||
Db 109709 GAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCCGCA 109748


RESULT 7
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag=  a
```

```
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             42.6%;  Score 43;  DB 22;  Length 20394;
  Best Local Similarity   64.6%;  Pred. No. 0.0019;
  Matches   64;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       3 TCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAATG 62
           ||||   ||| ||||||  |  || ||||    || |||  ||||    ||  ||| | |
Db   14738 TCGCGACGGTCCTGGGCCACGCCGGCCCGGAAACGATCGCCCCTGACCGGGCCTTCCAGG 14797

Qy      63 AGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           | |  ||||||||||||  ||| |||| | ||  |||| |
Db   14798 ACCTCGGCCTCGACTCCCTGACGGCGATCGAACTCCGTA 14836


RESULT 8
AAV21186
ID   AAV21186 standard; DNA; 5676 BP.
XX
AC   AAV21186;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..5676
FT                   /*tag=  a
FT                   /product= "polyketide synthase"
FT                   /note= "no stop codon given"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52844.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 38-43; 205pp; English.
XX
CC   The present sequence represents Amycolatopsis mediterranei strain wt3136
CC   5.7 kb KpnI fragment DNA, from the present invention. The present
CC   invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC   gene cluster DNA fragment comprising a DNA region involved directly or
CC   indirectly in the gene cluster responsible for rifamycin synthesis,
CC   including the adjacent DNA regions to the right and left which, by
CC   reason of their function in connection with rifamycin biosynthesis,
CC   qualify as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 19;  Length 5676;
  Best Local Similarity   64.0%;  Pred. No. 0.0025;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           ||||| | |||||| ||   |  ||  |||   ||  ||    | |  | || ||||||
Db    3132 GTCGCGCTGGTGCTCGGGCACGCCGGGCCGGAGGCCGTCCGCGCGGACACGGCGTTCAAG 3191

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||    ||| |||||||| ||| | |||||||  | ||||
Db    3192 GACACCGGCTTCGACTCGCTGACGTCGGTGGAACTGCGCA 3231


RESULT 9
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag=  b
```

```
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 19;  Length 53789;
  Best Local Similarity   64.0%;  Pred. No. 0.0033;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           |||||    ||||| ||   |  |||| |||||||  ||      |  | ||  |||||
Db   25291 GTCGCGGCCGTGCTCGGGCACGCCGACGCGAGCGCCGTCCGCGTCGACACGGCCTTCAAG 25350

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||    ||| |||||||| |||  ||||| ||| | ||||
Db   25351 GACGCCGGCTTCGACTCGCTGACCGCGGTCGAGCTGCGCA 25390
```

```
RESULT 10
AAZ87298
ID   AAZ87298 standard; DNA; 11220 BP.
XX
AC   AAZ87298;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..111220
FT                   /*tag= a
FT                   /product= "PikAII"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77193.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 398-403; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;
```

```
  Query Match             41.4%;  Score 41.8;  DB 21;  Length 11220;
  Best Local Similarity   63.4%;  Pred. No. 0.0041;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           |.||||   ||| ||   |  || | |||    ||  ||||  | |||  ||  |||
Db    4362 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 4421

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db    4422 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 4462


RESULT 11
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT   CDS             14056..14136
FT                   /*tag=  b
FT   CDS             14148..15827
FT                   /*tag=  c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             41.4%;  Score 41.8;  DB 18;  Length 15872;
  Best Local Similarity   63.4%;  Pred. No. 0.0043;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||    || || |  | |  |||||||| |    |||| |  ||  |    |||||
Db    2833 CGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAA 2892

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||||  |||| |||||||  | ||| | || ||| | || |
Db    2893 GGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGA 2933


RESULT 12
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag=  b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag=  c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
```

```
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match             41.4%;  Score 41.8;  DB 21;  Length 15872;
  Best Local Similarity   63.4%;  Pred. No. 0.0043;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||    || || |  | |  |||||||| |    |||| |  ||  |    |||||
Db    2833 CGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAA 2892

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
             |||| |||| |||||||  | ||| | || ||| | || |
Db    2893 GGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGA 2933


RESULT 13
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 31.
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

  Query Match             41.4%;  Score 41.8;  DB 21;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 0.0047;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db   20049 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 20108

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           ||| |  ||  |||||||| ||| |||||| ||| ||||||
Db   20109 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 20149


RESULT 14
AAZ87285
ID   AAZ87285 standard; DNA; 37948 BP.
XX
AC   AAZ87285;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
```

```
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77180.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 13; Page 299-315; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

  Query Match             41.4%;  Score 41.8;  DB 21;  Length 37948;
  Best Local Similarity   63.4%;  Pred. No. 0.0047;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db   21219 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 21278

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||  |  ||  |||||||| ||| |||||| ||| ||||||
Db   21279 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 21319


RESULT 15
AAA75633
ID   AAA75633 standard; DNA; 38506 BP.
XX
AC   AAA75633;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 15-32; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             41.4%;  Score 41.8;  DB 21;  Length 38506;
  Best Local Similarity   63.4%;  Pred. No. 0.0047;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           | ||||   ||| ||   |  || |  |||    ||  ||||  | |||  ||  |||
Db   18191 CCTCGCCGTGGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCG 18250

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||  |  ||  |||||||| ||| |||||| ||| ||||||
Db   18251 TGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCCGCA 18291


Search completed: November  5, 2002, 13:10:44
Job time : 264.215 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_24950_25050
Perfect score: 101
Sequence: 1 cgtcgctcgggtgctgggct.....gatggcggtggagatccgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	68750	1	AF210843	AF210843 Sorangium
3		101	100.0	68750	6	AR193029	AR193029 Sequence
4		101	100.0	68750	6	AR199551	AR199551 Sequence
5		101	100.0	68750	6	AR199559	AR199559 Sequence
6		101	100.0	68750	6	AR199567	AR199567 Sequence
7		101	100.0	68750	6	AR201097	AR201097 Sequence
8		101	100.0	68750	6	AR208671	AR208671 Sequence
9		101	100.0	71989	6	AR172664	AR172664 Sequence
10		65.8	65.1	10910	1	AX024383	AX024383 Sequence
11		65.8	65.1	10910	6	AX024276	AX024276 Sequence
12		48.8	48.3	43280	1	SFU78289	U78289 Streptomyce
13		47.2	46.7	33529	6	AR166425	AR166425 Sequence
c 14		45.6	45.1	24225	1	SC2C4	AL512902 Streptomy
15		45.6	45.1	65140	6	AX211705	AX211705 Sequence
16		45.6	45.1	104326	1	AB070940	AB070940 Streptomy
17		45.6	45.1	123580	1	AF263912	AF263912 Streptomy
18		45.6	45.1	125401	6	AX211739	AX211739 Sequence
19		45	44.6	28732	1	AF082100	AF082100 Streptomy
20		44	43.6	32870	1	AF007101	AF007101 Streptomy
c 21		43.4	43.0	15970	1	AE007061	AE007061 Mycobacte
c 22		43.4	43.0	53450	1	MTV018	AL021899 Mycobacte
c 23		43.4	43.0	107379	1	SHGCPIR	X86780 S.hygroscop
24		43	42.6	20394	1	SNA132222	AJ132222 Streptomy
25		43	42.6	20394	6	AX067996	AX067996 Sequence
c 26		43	42.6	84985	1	SNA278573	AJ278573 Streptomy
c 27		42.8	42.4	49736	1	AF319998	AF319998 Stigmatel
28		42.4	42.0	5676	6	A69718	A69718 Sequence 1
29		42.4	42.0	53784	1	AMM223012	AJ223012 Amycolato
30		42.4	42.0	53789	6	A69720	A69720 Sequence 3
31		42.4	42.0	90445	1	AF040570	AF040570 Amycolato
c 32		42	41.6	82746	1	AF453501	AF453501 Actinosyn
33		41.8	41.4	30000	6	AX250261	AX250261 Sequence
34		41.8	41.4	30000	6	AX250262	AX250262 Sequence
35		41.8	41.4	37948	1	AF079138	AF079138 Streptomy
36		41.8	41.4	78210	1	AB070949	AB070949 Streptomy
37		41.8	41.4	107379	1	SHGCPIR	X86780 S.hygroscop
38		41.4	41.0	9513	6	AX089460	AX089460 Sequence
39		41.4	41.0	36181	1	U00023	U00023 Mycobacteri
40		41.4	41.0	41782	1	AY118081	AY118081 Streptomy
41		41.4	41.0	50000	6	AX089417	AX089417 Sequence
42		41.4	41.0	50000	6	AX089420	AX089420 Sequence
43		41.4	41.0	80161	1	AY007564	AY007564 Saccharop
44		41.4	41.0	80161	6	AR165018	AR165018 Sequence
c 45		41.4	41.0	308950	1	MLEPRTN9	AL583925 Mycobacte

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-643 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"

CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRSERHLPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"

CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLGPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFEPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"

gene 1998. .6263
/gene="epoA"

CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRPERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLPEVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"

gene 6260. .10492
/gene="epoB"

CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"

gene 10639. .16137
/gene="epoC"

CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRILLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRRIEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24950 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 25009

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25010 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 25050


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
CDS                  6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene                 7610. .11875
                     /gene="epoA"
CDS                  7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADEPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSFRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVGLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPAEVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature         7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
misc_feature         9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
misc_feature         10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
misc_feature         11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
gene                 11872. .16104
                     /gene="epoP"
CDS                  11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene                 16251. .21749
                     /gene="epoB"
CDS                  16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match             100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy      61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
```

```
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  30562 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 30621

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||| ||||||||||
Db  30622 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 30662


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24950 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 25009

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25010 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 25050


RESULT 10
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           65.1%;  Score 65.8;  DB 1;  Length 10910;
  Best Local Similarity   78.2%;  Pred. No. 1.1e-06;
  Matches   79;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          |||||| ||||||||||||||||  ||||||| ||||||||||    ||    || |||
Db   2000 CGTCGCCCGGGTGCTGGGCTTCTCCGACCCGGGCGCGCTCGACGTCGGCCGAGGCTTCGC 2059

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          ||||||||  ||||||||  |||||||  |||||||||| |
Db   2060 CGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATCCGTA 2100


RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           65.1%;  Score 65.8;  DB 6;  Length 10910;
  Best Local Similarity   78.2%;  Pred. No. 1.1e-06;
  Matches   79;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy      1 CGTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAA 60
          |||||| ||||||||||||||||  ||||||| ||||||||||    ||    || |||
Db   2000 CGTCGCCCGGGTGCTGGGCTTCTCCGACCCGGGCGCGCTCGACGTCGGCCGAGGCTTCGC 2059

Qy     61 TGAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          ||||||||  ||||||||  |||||||  |||||||||| |
Db   2060 CGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATCCGTA 2100


RESULT 12
SFU78289
LOCUS       SFU78289               43280 bp    DNA     linear   BCT 13-AUG-1997
DEFINITION  Streptomyces fradiae tylactone synthase, starter module and modules
            1-7, (tylG) gene, complete cds.
ACCESSION   U78289
```

```
VERSION     U78289.1  GI:2317859
KEYWORDS    .
SOURCE      Streptomyces fradiae.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Sequence of Streptomyces fradiae tylactone synthase gene tylG
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
            Center, Indianapolis, IN 46285, USA
COMMENT     tylG.
FEATURES             Location/Qualifiers
     source          1. .43280
                     /organism="Streptomyces fradiae"
                     /db_xref="taxon:1906"
     gene            1. .43280
                     /gene="tylG"
     CDS             816. .14234
                     /gene="tylG"
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tylactone synthase starter module and modules 1
                     & 2"
                     /protein_id="AAB66504.1"
                     /db_xref="GI:2317860"
                     /translation="MSSALRRAVQSNCGYGDLMTSNTAAQNTGDQEDVDGPDSTHGGE
                     IAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWRAALEDHAGFDAGFFGMNAR
                     QAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYAVLTRRSAVSAG
                     GYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGETSLAVAG
                     GVNLILTEESTTVMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAALADGD
                     RVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTGTRA
                     GDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGEL
                     VPSLNFSTPNPDIPLDDLRLRVQTERQEWNEEDDRPRVAGVSSFGMGGTNVHLVIAEA
                     PAAAGSSGAGGSGAGSGAGISAVSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLA
                     DVAVTMADRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVLDGGVVVGAAP
                     GGAGAAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSV
                     HVGWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAA
                     TVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVN
                     GPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRP
                     RSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHP
                     VLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL
                     TTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR
                     HTAVVLGLDTAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTPRA
                     LAQFLHTELLDGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSPEALWRLVVEG
                     VDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGISPREAVAM
                     DPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRLAEGAEGSDGYLLT
                     GTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMALAGGVTVM
                     AGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAV
                     VCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGDPIE
                     AQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVD
                     EPSRHVDWSAGGVWLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAPDTAEAESA
                     TTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVT
                     MAGRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVF
                     PGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVSGGAGLERVDVVQP
                     VTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGR
                     YLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVNGPASTVVSGDRRAVAGYVAVCQAE
                     GVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAG
                     YWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLR
                     RQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHLTTLPTYPFNHHHYWLDTTPTTPAT
                     TTQSPTDAWRYRVTWKALTESSPVRPHSIGRCLLVAPPTTDGELLDGLTTVLSERGAS
                     VARLEVPIGARRAEVAELLKPSMESAGEENTTVVSLLGLVPSTDAVRTSIALLQAVSD
                     IGVPAARVWALTRRAVAVVPGETPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAEL
                     TRTPETSQPPQTPERLPQTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAA
                     AGAASWQPSGTVLITGGMGAIGRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELR
                     GHGCEVVHAACDVAERDALAALVTAYPPNAVFHTAGILDDAVIDTLSPESFETVRGAK
                     VCGAELLHQLTADIKGLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLP
                     ATSVAWGLWGGGGMAAGAGEESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVDW
                     ERFAPATNAIRPGRLFDTVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKL
                     VELVRTEAAFVLRHPNTDAIGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFD
                     HPSPTALARLLLDRLTGAGAPAPAADEPPLPVAVADDDPVVIVGMACRFPGGAGTPEA
                     LWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGI
                     SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPE
                     EYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMA
                     LAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARR
                     LGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVDVVEAHGTG
                     TRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGW
                     LPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAP
                     EALEALEATDAPEAPEAPEAPDVTDVTEALEAPDATEAEGAKAPGSPEEAQPAVGVVP
                     VVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVTMAGRSRFGYRAVVLARGEAELAG
                     RLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAA
                     SMRECARALSVHVEWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAV
                     VGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA
                     KWPGVQVAAVNGPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKG
                     ELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEQG
                     HRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATL
                     TWDPALPPGHLTTLPTYPFNHHHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLA
                     DSGERVFAGRLAGSEHDWLTDHAVSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEA
                     PLVLPAAGGVRVQMRVSAADESGRRRVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL
                     VPDPEPTPPDADWARAWPPAGERVEPAELYERFGALGYEYGEAFAGVRAVWRQPDALL
                     AEVLLPDRASTGAGRFGVHPALLDAALQPWIAGGLLEVPEDAVLLPFAWQGVSLYATG
                     AGALRVRLTKAGDGAVSLQAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGERSLY
                     RVEWQPRLLPAGPPRSWAVLGPDADRLAGTPGLGDQPDGGPTALYPEVRALRKALAAG
                     APRPEAVVLPVLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAV
                     PGEEIGDLACAGVWGLVRSARSEHPGRFALVDTDGHPDDRTALPLALRAVLDGAGQLS
                     LRAGTARTPVLLRAGTPEEQRGPAFDPAGTVLVTGATGTLGRLLARHLAAEHGVRHLL
                     LLSRGGRAAEGADELAAELAGLEAEPCFAACDAADREALARVLAEVPADRPLTGVIHA
                     AGVLDDGTLDALTPERIGTVMRPKADAALNLHELTRTSPLSVFAVFSGAAGILGRPGQ
                     ANYAAANTFLDALAQHRRAHGLPAVSLAWGLWGGATGMTGHLSGTDLRRMRRSGIAPM
                     THDQGLALFDRALAASAEDPLLVPMRLDLAALVRERAEHGPDAVPGPLLGLLPARAAV
                     RQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEP
                     GRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRDGG
                     DTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAALTPQGRNPSAPAPAPSDL
                     DERLDSANDDDLFAFIEEQL"
     misc_feature    942. .3929
                     /gene="tylG"
                     /function="encodes starter module"
     misc_feature    942. .2156
                     /gene="tylG"
                     /function="encodes beta-ketoacyl ACP synthase KS-Q of
                     starter module"
     misc_feature    2571. .3557
                     /gene="tylG"
                     /function="encodes acyltransferase domain of starter
                     module"
     misc_feature    3675. .3929
                     /gene="tylG"
                     /function="encodes acyl carrier protein domain of starter
                     module"
     misc_feature    3993. .8471
                     /gene="tylG"
                     /function="encodes module 1"
     misc_feature    3993. .5264
                     /gene="tylG"
                     /function="encodes beta-ketoacyl ACP synthase of module 1"
     misc_feature    5631. .6617
                     /gene="tylG"
                     /function="encodes acyltransferase domain of module 1"
     misc_feature    7410. .7949
                     /gene="tylG"
                     /function="encodes beta-ketoreductase domain of module 1"
     misc_feature    8220. .8471
                     /gene="tylG"
                     /function="encodes acyl carrier protein domain of module
                     1"
     misc_feature    8541. .13970
                     /gene="tylG"
                     /function="encodes module 2"
     misc_feature    8541. .9812
                     /gene="tylG"
                     /function="encodes beta-ketoacyl ACP synthase of module 2"
```

```
     misc_feature    10260. .11246
                     /gene="tylG"
                     /function="encodes acyltransferase domain of module 2"
     misc_feature    11319. .11876
                     /gene="tylG"
                     /function="encodes dehydratase domain of module 2"
     misc_feature    12861. .13415
                     /gene="tylG"
                     /function="encodes beta-ketoreductase domain of module 2"
     misc_feature    13719. .13970
                     /gene="tylG"
                     /function="encodes acyl carrier protein domain of module
                     2"
     CDS             14351. .19945
                     /gene="tylG"
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="tylactone synthase module 3"
                     /protein_id="AAB66505.1"
                     /db_xref="GI:2317861"
                     /translation="MTAELVATRKRLGALEERAREPIAVVAMSCRYPGGVTTPEDLWR
                     LLADERDAVSGLPRDRGWDLDALYDPDGGPGTSYAREGGFLSHCAGFDAEFFGISPRE
                     ALAMDPQQRLLLETSWEALERAGVTADRARGSRTGVYAGVMYDDYGARVLYGAGAGPP
                     EDLEGYLVNGSAGSIASGRVSYTFGLRGPAVTVNTACSSSLVSLHLAVRALRNGECDM
                     ALAGGATVLSTPTVLVDFSRQRGLAPDGRCKAFADSADGTSWAEGAGMLLLQRLSDAR
                     REGRPVLAVIRGSAVNQDGASNGLTAPNGRAQRQVIEDALRDAGVGPDQVDAVEAHGT
                     GTELGDPIEAGALLATYGQARTAERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMRHG
                     RLPRTLHVDRPTTRVDWEKGGVRLLTEPVPWPGEAGEPRRAGVSSFGASGTNAHVVLE
                     SVPAGEPPAAGRPEDTGGAWTVSGRGPAALRAQAARLYDALTGTGTGTGQGAGQGAGP
                     GTAEVAGALAHARTAFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERR
                     TAFLFSGQGSQRAGSGRGLYRRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEA
                     EPLDRTEFTQPALFALQTALFRLAEHHGLRAEALCGHSVGEIAAAHAAGVLTLPDAAR
                     LVAARGRLMQALPAGGAMVALRATAEEIAPLLERRAGELALAAVNGPSSVVVSGDEAA
                     VLELLEQWRAEGREARRLRVSHAFHSPRMDGMLTQFDRVARTLTFAPPTIPLVSTLTG
                     TPVTEETLCTADHWVRQARGEPVRFLDAMRTLRADGIDTFVELGPDGVLSAMARDCADD
                     RPDGDTTGAGDGETPDPLLPTLPLLRRSVPETGDAEHPGGFERALATAYAHGVPLRLAP
                     APDAASLAVAAELPTYAFQRTHYWLDAPAAPAALPAGLDDAGHPLLSAALDLPGGRGT
                     VWTGALSAATLPWAADHSVHGRTVLPGTALLDLALHAAPRVGELTFEAPLVLPEDGEV
                     RLRVVLAEPDASGVRELSVHSAGEDGGWTRHATAVLDTGTTTAGEPAGAPPAAWPPGD
                     AEPLDLAAEYERFADAGIGYGPAFRGLRSAWRDGDAILADVRLPGELAGEADRYGIHP
                     ALLDAALHTAAAALGGAHGMLPFTWNGVTLHARGAHAIRVRLTPAGPDAVAVTAVDPA
                     GRPVFTAASLTLRPVTTGQLTAAEAARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDL
                     LPGETPHHPDLASALADTGTAPFRVLADLRGYGTATPRELASQALALVQQWADAAEAA
                     EGRLVLVTRRAVDIGDGVTDPAAATVWGLVRAAQSEHPGCFALLDTDDSPRSRQLLPR
                     VAGTAEQLALRDGTLLAPSLTRATLPAGARLPALDGTVLITGGTGSLGAEAARHLVTR
                     HGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAACDVADRAALRALLDRVPAGHP
                     LTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLHELTQGHALSAFILYSSAAG
                     VLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPWDGGGMASGLGGTDTARLRR
                     SGIAPLSRAEGLAALDAALAAGGDDTAPAHLLPIRVDAVTLRGADTVPAVLRDLAGTA

  Query Match             48.3%;  Score 48.8;  DB 1;  Length 43280;
  Best Local Similarity   68.0%;  Pred. No. 0.014;
  Matches   68;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
           ||  |  | ||||||||||.|   | ||||  |||| | || || |||     ||||||
Db   13779 GTGTCGGGCGTGCTGGGCTACTCGGGCCCGCACGCGGTGGAGCCCGGGCGCCCGTTCAAG 13838

Qy      62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGCAGATCCGCA 101
           || |  ||| |||||||| |||  || ||||||  | ||||
Db   13839 GATCTCGGCTTCGACTCGCTGACCGCCGTGCAGCTGCGCA 13878


RESULT 13
AR166425
LOCUS       AR166425               33529 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 3 from patent US 6280999.
ACCESSION   AR166425
VERSION     AR166425.1  GI:16241741
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 33529)
  AUTHORS   Gustafsson,C., Betlach,M.C., Ashley,G., Julien,B. and Ziermann,R.
  TITLE     Sorangium polyketide synthases and encoding DNA therefor
  JOURNAL   Patent: US 6280999-A 3 28-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .33529
                     /organism="unknown"
BASE COUNT     4489 a    9518 c   14470 g    5046 t      6 others
ORIGIN

  Query Match             46.7%;  Score 47.2;  DB 6;  Length 33529;
  Best Local Similarity   69.6%;  Pred. No. 0.036;
  Matches   64;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAATGAG 64
           ||   |||||||||| |    || ||||||| ||| || || | |   ||.||   ||
Db    9459 GCGGCGGTGCTGGGCGTGAAGGATCCGAGCGGGCTGGACCCGGAGCGAGGCTTTCTGGAC 9518

Qy      65 CAGGGCCTCGACTCGTTGATGGCGGTGGAGAT 96
           |  || || |||||||||||||||||||| |
Db    9519 CTCGGGCTGGACTCGTTGATGGCGGTGGAGCT 9550


RESULT 14
SC2C4/c
LOCUS       SC2C4                  24225 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 2C4.
ACCESSION   AL512902 AL645882
VERSION     AL512902.2  GI:20520683
KEYWORDS    accA1; secreted FAD-binding protein; type I polyketide synthase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24225)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
REFERENCE   2  (bases 1 to 24225)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 24225)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:12274798.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
```

Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/ jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2C4.

FEATURES Location/Qualifiers

source 1. .24225
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"

misc_feature complement(2. .1477)
/note="previously sequenced DNA fragment. EMBL:AF113603 Streptomyces coelicolor putative acyl-CoA carboxylase complex A subunit (accA1) gene, complete cds."

misc_feature 2. .119
/note="nominal overlap with Streptomyces coelicolor cosmid SCAH10"

misc_feature complement(374. .1471)
/note="Pfam match to entry PF00289 CPSase_L_chain, Carbamoyl-phosphate synthase (CPSase), score 585.30, E-value 3.8e-172"

misc_feature complement(602. .625)
/note="PS00867 Carbamoyl-phosphate synthase subdomain signature 2"

RBS complement(1485. .1489)

RBS 1774. .1778

gene 1786. .3438
/gene="SC2C4.02"
/note="SCO6272"

CDS 1786. .3438
/gene="SC2C4.02"
/note="SC2C4.02, possible secreted FAD-binding protein, len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222) Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores: opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa overlap. Contains Pfam match to entry PF01565 FAD_binding_4, FAD binding domain. Also contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted FAD-binding protein"
/protein_id="CAC22143.1"
/db_xref="GI:12274800"
/db_xref="SPTREMBL:Q9EX55"
/translation="MTEVSRRKLMKGAAVSGGALALPALGAPPATAAPAAGPEDLPGP
AAAAAGRRPAPAPVHVTRSDPRYQGLISGSNQRWVGNPDYIRVVSSAGQVVRAVQEAV
DKGLKIAVRSGGHCDEDFVANRDVRVVIDMAGMDSVTYDRERRAFAVGPGARLGTVYR
TLYKRWGVVLPGGTCPTVGAGGHITGGGYGALSRSRGLTVDHLYAVEVVHVDARGRAR
KVVATREEDDPNRELWWAHTGAGGGNFGVITRYWLRSPDAVGDDPSTLLPAAPSEVLL
SDVSWSWDDLDEASFTRLLRNFTEWHARNSAPDSPGRFLFSQLKTMHKAAGYFRMSTQ
VDAAAPDADRLLDDYLAAISEGTGVTYHVGDRYRAPWLYAVTEWSGFVEASVPRWKSK
SAYVREVMPEEQLRAVYRQLTRDDYPGPYGMIAIVGFGGKINEVAPGDTATAQRDSIA
KMLYCSLWSDPADDALHQRWIREAYEDVYASTGGVPRPGGVNDGCYINYADADLADPA
LNRSGIPWHELYFKGNYPRLQRVKATWDPRNVFSHRLGIRKP"

misc_feature 1948. .2535
/gene="SC2C4.02"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD binding domain, score 6.50, E-value 1.1e-06"

gene complement(3514. .9972)
/gene="SC2C4.03c"
/note="SCO6273"

CDS complement(3514. .9972)
/gene="SC2C4.03c"
/note="SC2C4.03c, probable type I polyketide synthase, len: 2152 aa; N-terminal domain almost identical to TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2) type I polyketide synthase (fragment), 720 aa; fasta scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity in 721 aa overlap and C-terminal domain similar to SW:HETM_ANASP (EMBL:L22883) Anabaena sp. polyketide synthase HetM, 506 aa; blastp scores: Score= 761 (267.9 bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identities= 184/491 (37%), Positives= 278/491 (56%). Also highly similar, in this same cosmid, to the N-terminal region of SC2C4.04c and to SC2C4.05c. Contains Pfam matches to entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase, PF00698 Acyl_transf, Acyl transferase domain and PF00550 pp-binding, Phosphopantetheine attachment site and matches to Prosite entries PS00606 Beta-ketoacyl synthases active site, PS00340 Growth factor and cytokines receptors family signature 2 and PS00012 Phosphopantetheine attachment site"
/codon_start=1
/transl_table=11
/product="putative type I polyketide synthase"
/protein_id="CAC22144.1"
/db_xref="GI:12274801"
/db_xref="SPTREMBL:Q9EX54"
/translation="MSNEQKMVEYLKWVTTDLQKARQRLAELEAGNDEPVAIVGMACR
FPGGVASPDDLWRLVMDEADGITDWPTDRGWDIDAIYDPERGKPGKSYTRKGGFLDGA
AQFDPGFFGISPREALAMDPQQRVLLETAWETFEQAGIDPATLRGSQVGVFAGVVEQS
YLGLEGPEELEGYLMTSKLSSVASGRIAYTFGFEGPAVSVDTACSSSLVALHLAVQSV
RKGESTMALAGGSTVSGNPSGFVDFSRQNGLAEDGRIKSFAAAADGTAWSEGVGLLLV
EKLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQERVIHQALAGARLNASDVD
VVEAHGTGTRLGDPIEAQALLASYGRQRPADRPLYLGSLKSNIGHTVAAAGAGGVIKM
IMAMRHGILPKTLNVDEPTPMVDWEAGAVELLTEARPWPETGAPRRAGVSAFGVSGTN
AHVILEEPPAEEPEEPADTAVTALPVLPWALSARTAPALRDQARRLLSHVEERPAEDP
LDVAYSLAAERSALEHRAVVVGSDREELLTGLRALAEGRTTPTGAVQDTRQTGKTAFL
FTGQGAQRTGMGMDLYRTFPAYAHAFDTITAQLDPHLDQPLHHTITTGHHLHHTGNTQ
PALFATEVALYRLLETWGITPDYLAGHSIGELTAAHISGILTLQDACTLVTARARLMQ
NLPTHGTMIALQATEQEILPHLTGHEHHLTIAAINSPTSLVISGNQTAAQHIATQLQA
QGRKTKTLTVSHAFHSPHMDGMLHDFHHTATQLTYHEPTIPIVSTLTGNLATHNDLRT
PTYWTDQLRNTVRYTQAIHTLHTAGVTTYTEIGPDATLTPLTQQTLDTTTAIPLLRHN
QPETHTLLTAISRLHNRGIPINWKNLFTTTPAKHTPLPTYAFQHERFWVETSGVAADA
SDLGLESAGHPMLGAAVTVAGADQVLFTSRLSLRTHPWLADHVVLDQTLVPATALVEM
AIRAGDELGSTVLNDLSVLGPLVLPGKGGVHVQVSVGSPDASGLRELSIYSRPDNADV
PWTLNARGRLSFRGPDAPFSLAQWPPAGAEEVPLDEAYDKLAVQGYVYGPVYRGLEHV
WRLGDEVYADVRLPDGSRTGIGGFGLHPALLETALAASCLAGYGSSPEGAVWISTDWK
DVRLHATGADRLRVRLAPDADGALNVQLADRAGRPVASVSSVACRAIAVGEVSNALAR
NRDSLFHVDWAPFPVAPSAGTERWVSLGADEPGARRYPDVASLARAVAAGESVDVALV
GTETGDTGDRVADVHDVTREALALVQDWLAEERLAGTRLVVRTRGGVTTAGDEDVTDL
GAAAVWGLLRSAQSENPGRIVLVDVDDEEASARILPAVVASGEPQSAVRDGQVFVPRL
GRVPEAPAGAATGAWDRRGTVLITGGTGSLGGLFARHLVVEHGVRNLLLTSRRGRDAE
GAAELEAELTALGARVEIAACDAADRAALAAVLADIPAEYPLTGVLHTAGVLDDGLIG
SQTPQRLSAVLRPKVDATWNLHELTRDMDLSAFVLFSSIAAVVGGPGQSTYSAANHFL
DALAEHRRSRGLAATSLAWGLWSQTSGMTGHLDEADLKRIARSGFRPVESEQGPALLD
IALTVGRAGLVATPMDLTALREQVQVPVLLSKLARTPVRRTAQNAEVSSESLAERLAG
LGDAEQREAVLGIVLGQAAVVLGHSDATGIDRQRSFTALGFDSLTSVELRNQLGTSTG
LRLPPTLVFDHPTPVALAEYLHGEVRAELSADGSASRPGVDFAAEVRLADDVRPADEV
ITTAADPKEILLTGASGFLGAFLLRDLMRTTTARVHCLVRGADEAAAMERLKANAEWY
RVWDDIDPERVSVVLGDLAEPRLGLDEESFDALARTVDVVYHNGARVHWLLPYETLKA
SNVTGTEEVLRLAARHRTVPVHYVSTVGVFDGVREPGVPLRVTDPTGPAESLPSGYLR
SKWVAEQVIEVARDRGLPVSVYRVDVISGDRVNGACQTRDFVWLTLKGLIQARSVPKG
TEGRFHLLPVDYVSAAITGISRQPGTVGRTFHLFNQSSLALSQCVELLRSLGYELDEV
DWDTWTQVVTSGDNALLPLLDAFEMMTSDTDGFYPPIDTAETVAALEGTGIGIPVLTR
ELFEKYVAFFVEEGHFPAAG"

misc_feature complement(4732. .4935)
/gene="SC2C4.03c"
/note="Pfam match to entry PF00550 pp-binding, Phosphopantetheine attachment site, score 68.50, E-value 2.5e-17"

misc_feature complement(4810. .4857)
/gene="SC2C4.03c"
/note="PS00012 Phosphopantetheine attachment site"

misc_feature complement(5248. .5268)
/gene="SC2C4.03c"
/note="PS00340 Growth factor and cytokines receptors

```
                    family signature 2"
     misc_feature   complement(7342. .8286)
                    /gene="SC2C4.03c"
                    /note="Pfam match to entry PF00698 Acyl_transf, Acyl
                    transferase domain, score 248.50, E-value 9.5e-71"
     misc_feature   complement(8596. .9867)
                    /gene="SC2C4.03c"

  Query Match            45.1%;  Score 45.6;  DB 1;  Length 24225;
  Best Local Similarity  66.0%;  Pred. No. 0.092;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
          || ||   |||||| |||  |  ||| ||  || |  |||| || ||   ||  ||| |
Db  21339 GTGGCCGCGGTGCTCGGCCACTCCGATCCCGGCTCCGTCGACCCCGGCCGGGCCTTCCAG 21280

Qy     62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||| |||| |||||||  |||  || || ||  ||||||
Db  21279 GAGCTGGGCTTCGACTCACTGACAGCCGTCGAACTCCGCA 21240


RESULT 15
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES            Location/Qualifiers
     source         1. .65140
                    /organism="Streptomyces noursei"
                    /db_xref="taxon:1971"
                    /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match            45.1%;  Score 45.6;  DB 6;  Length 65140;
  Best Local Similarity  66.0%;  Pred. No. 0.081;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      2 GTCGCTCGGGTGCTGGGCTTCCTCGACCCGAGCGCGCTCGATCCTGGGATGGGGTTCAAT 61
          | |||   |||||| ||   |  ||||      ||| ||| | | |||     |||| |
Db  49388 GCCGCGGTGGTGCTCGGGCACGCCGACGGCTCGGCGATCGGTGCGGGGCGACAGTTCCAG 49447

Qy     62 GAGCAGGGCCTCGACTCGTTGATGGCGGTGGAGATCCGCA 101
          |||  |||| |||||||| |||  |||||||||| ||||||
Db  49448 GAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCCGCA 49487


Search completed: November  5, 2002, 12:24:02
Job time : 603.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400
Perfect score: 101
Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactggcggagacgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	36.2	35.8	1084	17	AG077752	AG077752 Pan trogl
	2	34	33.7	1296	17	AG073215	AG073215 Pan trogl
	3	32.4	32.1	905	17	AQ688230	AQ688230 nbxb0077G
c	4	32.2	31.9	694	12	BF422350	BF422350 Lr_ad_142
	5	31.8	31.5	300	13	BJ248635	BJ248635 BJ248635
	6	31.4	31.1	572	13	BI875065	BI875065 963121E12
	7	31	30.7	484	13	BI336276	BI336276 AR094C04S
	8	31	30.7	547	12	BF078123	BF078123 228383 MA
c	9	30.8	30.5	413	10	AV630710	AV630710 AV630710
	10	30.6	30.3	479	14	BQ294770	BQ294770 WHE2854_B
c	11	30.6	30.3	523	10	AW566146	AW566146 660062F05
c	12	30.6	30.3	590	17	BH782972	BH782972 fzmb011f0
	13	30.6	30.3	610	9	AI621801	AI621801 486026B10
	14	30.6	30.3	879	12	BG755515	BG755515 602713891
	15	30.6	30.3	1024	11	AY107409	AY107409 Zea mays
	16	30.4	30.1	253	13	BI478653	BI478653 949068B09
c	17	30.4	30.1	747	14	BQ805689	BQ805689 WHE3569_H
	18	30.2	29.9	493	10	BE363468	BE363468 WS1_63_B0
	19	30.2	29.9	540	13	BI809588	BI809588 F013D01 O
c	20	30.2	29.9	791	17	AG049188	AG049188 Pan trogl
c	21	30.2	29.9	1285	17	AG058001	AG058001 Pan trogl
c	22	30	29.7	270	9	AA215413	AA215413 zr95h04.s
c	23	30	29.7	446	9	AA279201	AA279201 zs83d11.s
c	24	30	29.7	457	9	AA828455	AA828455 oc42e01.s
c	25	30	29.7	501	9	AA827645	AA827645 od55h06.s
	26	29.8	29.5	796	17	AZ192927	AZ192927 SP_1022_B
c	27	29.8	29.5	823	12	BF984458	BF984458 602307808
	28	29.8	29.5	1421	17	AQ012194	AQ012194 430P1A083
	29	29.6	29.3	464	14	BQ159769	BQ159769 WHE2230_F
c	30	29.6	29.3	542	12	BG274058	BG274058 WHE2231_A
c	31	29.6	29.3	547	12	BG274590	BG274590 WHE2235_F
c	32	29.6	29.3	558	12	BG274373	BG274373 WHE2227_H
c	33	29.6	29.3	558	12	BF292551	BF292551 WHE2215_B
c	34	29.6	29.3	565	14	BQ841584	BQ841584 WHE4212_G
c	35	29.6	29.3	587	12	BF291659	BF291659 WHE2206_D
c	36	29.6	29.3	598	12	BF291479	BF291479 WHE2223_D
c	37	29.6	29.3	667	14	BQ841258	BQ841258 WHE4208_H
c	38	29.6	29.3	669	14	BQ841464	BQ841464 WHE4211_C
c	39	29.6	29.3	680	12	BF291310	BF291310 WHE2221_F
c	40	29.6	29.3	681	14	BQ840803	BQ840803 WHE4203_F
c	41	29.6	29.3	683	14	BQ841101	BQ841101 WHE4207_A
	42	29.6	29.3	742	12	BF616609	BF616609 HVSMEc001
c	43	29.6	29.3	743	14	BQ840751	BQ840751 WHE4203_B
c	44	29.6	29.3	764	14	BQ840589	BQ840589 WHE4201_C
	45	29.6	29.3	978	12	BG860535	BG860535 1024071C0

ALIGNMENTS

RESULT 1
AG077752/c
LOCUS AG077752 1084 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-072F16.R, genomic survey sequence.
ACCESSION AG077752
VERSION AG077752.1 GI:16629554
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-072F16.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1084)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

```
            PRIMERS
                Sequencing: M13Rev
            LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Qualifiers
     source         1. .1084
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-072F15.R"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      248 a    489 c    307 g     13 t     27 others
ORIGIN

  Query Match             35.8%;  Score 36.2;  DB 17;  Length 1084;
  Best Local Similarity   62.9%;  Pred. No. 4.4;
  Matches   56;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       8 CGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTG 67
           ||  | |  ||| ||||| || | || || |||| | | |||| |||   ||||||
Db     674 CGCCGTCTCCGCGGGTTTTCTCCGCCGCGCGCCCGGCGGGTGGGGGGGTCTTCGCCGCCC 615

Qy      68 GCGTCAGCGTCATGCGTCCACTGGCGGAG 96
           ||| |  ||   |||| ||   ||||| |
Db     614 GCGGCTTCGCGTTGCGCCCGGCGGCGGCG 586


RESULT 2
AG073215
LOCUS       AG073215               1296 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-064P04.F, genomic survey sequence.
ACCESSION   AG073215
VERSION     AG073215.1  GI:16625017
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-064P04.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1296)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES            Location/Qualifiers
     source         1. .1296
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598'
                    /clone="PTB-064P04.F'
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      338 a    287 c    531 g    126 t     14 others
ORIGIN

  Query Match             33.7%;  Score 34;  DB 17;  Length 1296;
  Best Local Similarity   58.6%;  Pred. No. 18;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           || || | || | |   || |   || |||| |||||| | || | |||| |  | || |
Db    1183 GTAGACCGGACGCCCGCGGCCGCGGTCTTCGGCCGAGCCGCGGCAGAGAGGAGGAGAGCGC 1242

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACG 99
           || | |||||| ||   || ||    |   | | |  ||
Db    1243 GCAGGTGGCGTGAGGTGCACGCCAGGAGNAGAGAACGCG 1281


RESULT 3
AQ688230
LOCUS       AQ688230               905 bp    DNA     linear   GSS 01-JUL-1999
DEFINITION  nbxb0077G03r CUGI Rice BAC Library Oryza sativa genomic clone
            nbxb0077G03r, DNA sequence.
ACCESSION   AQ688230
VERSION     AQ688230.1  GI:5329314
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 905)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GGAAACAGCTATGACCATG
            Class: BAC ends
            High quality sequence start: 46
            High quality sequence stop: 195.
FEATURES            Location/Qualifiers
     source         1. .905
                    /organism="Oryza sativa"
                    /strain="Japonica"
                    /cultivar="Nipponbare"
                    /db_xref="taxon:4530"
                    /clone="nbxb0077G03r"
                    /clone_lib="CUGI Rice BAC Library"
                    /tissue_type="Leaf"
                    /lab_host="E. coli DH10B"
                    /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                    HindIII; Rice is one of two most popular grains in the
                    world. Half of the world  population especially those
                    inhabiting highly populated areas of the  humid tropics
                    and subtropics, rely on rice as their primary source of
                    carbohydrate. Monocotyledonous rice is a diploid plant
                    (2n=24) with a  haploid genome equivalent of 431 Mbp
                    (Arumuganathan and Earle, 1991).  The relatively small
                    genome of rice, three times larger than that of
                    Arabidopsis, makes it suitable for genomic studies. In
                    order to  facilitate positional cloning, physical mapping
                    and genome sequencing  of rice, we have constructed a BAC
                    library from Oryza sativa, Nipponbare  variety. The
                    library contains 36,864 clones with an average insert size
                    of 128.5 Kb providing 10.9 haploid genome equivalents. The
                    deep  coverage allows the isolation a particular sequence
                    with a probability  of 99.9 %. Two high density filters,
                    each containing 18,432 clones  (doubly spotted), represent
                    the whole library for colony screening."
```

```
BASE COUNT      154 a    302 c    267 g    177 t      5 others
ORIGIN

  Query Match             32.1%;  Score 32.4;  DB 17;  Length 905;
  Best Local Similarity   62.2%;  Pred. No. 46;
  Matches   51;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      12 GACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGT 71
           | | ||  |||   |||| |||| || ||||  |||  | | ||||  |  | | | |||
Db      70 GGCGGCCGTGGCGGCGTCCGTCGTGCTCCCGGAGCGCCGAGAGGTGTCCCTCACCGCCGT 129

Qy      72 CAGCGTCATGCGTCCACTGGCG 93
           |  |||   ||||  ||| |||
Db     130 CCCCGTGGCGCGTGAACTTGCG 151


RESULT 4
BF422350/c
LOCUS       BF422350                694 bp    mRNA    linear   EST 28-NOV-2000
DEFINITION  Lr_ad_142_T3 Earthworm Lambda Zap Express Library Lumbricus
            rubellus cDNA clone Lr_ad_142 5' similar to gb|AAC53155.1| (U88623)
            aquaporin-4 - Mus musculus, mRNA sequence.
ACCESSION   BF422350
VERSION     BF422350.1  GI:11410339
KEYWORDS    EST.
SOURCE      humus earthworm.
  ORGANISM  Lumbricus rubellus
            Eukaryota; Metazoa; Annelida; Clitellata; Oligochaeta; Haplotaxida;
            Lumbricina; Lumbricidae; Lumbricus.
REFERENCE   1  (bases 1 to 694)
  AUTHORS   Sturzenbaum,S., Parkinson,J., Blaxter,M., Morgan,J., Kille,P.,
            Schaffner,W. and Georgiev,O.
  TITLE     Expressed Sequence Tags from the humus earthworm L. rubellus
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sturzenbaum SR
            School of Biosciences
            Cardiff University
            PO Box 911, Biomedical Building, Museum Avenue, Cardiff, CF10 3US,
            UK
            Tel: +44 2920 874119
            Fax: +44 2920 874094
            Email: SturzenbaumSR@cardiff.ac.uk
            Seq primer: T3 sequencing primer
            High quality sequence stop: 487.
FEATURES             Location/Qualifiers
     source          1. .694
                     /organism="Lumbricus rubellus"
                     /db_xref="taxon:35632"
                     /clone="Lr_ad_142"
                     /clone_lib="Earthworm Lambda Zap Express Library"
                     /tissue_type="whole worm"
                     /dev_stage="adult"
                     /note="Vector: pBK-CMV; The library was prepared using
                     protocol  given by supplier (Stratagene)."
BASE COUNT      159 a    172 c    211 g    152 t
ORIGIN

  Query Match             31.9%;  Score 32.2;  DB 12;  Length 694;
  Best Local Similarity   57.4%;  Pred. No. 51;
  Matches   58;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           |||  |||  | | |  || |  | || || |||  ||  |||   ||  | || || ||
Db     564 GTCAGACCTTTAAGAATGCCGCATCCGACGATCGCTCCTGCGCATTGACTGATGATGTAC 505

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
             ||| | | |||   || ||  ||  ||||||  || | |
Db     504 AGCGCAGCCCTCACGATCGTGATTCTTCTGGCGACGAGGAA 464


RESULT 5
BJ248635
LOCUS       BJ248635                300 bp    mRNA    linear   EST 05-APR-2002
DEFINITION  BJ248635 Y. Ogihara unpublished cDNA library, Wh_f Triticum
            aestivum cDNA clone whf7e20 5', mRNA sequence.
ACCESSION   BJ248635
VERSION     BJ248635.1  GI:20059371
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Ogihara,Y. and Murai,K.
  TITLE     Expressed genes in Triticum aestivum
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="whf7e20"
                     /clone_lib="Y. Ogihara unpublished cDNA library, Wh_f"
                     /tissue_type="spike at flowering date"
                     /dev_stage="Feekes' scale 10.5.1"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                     hydroponic conditions at UC Davis, salt stressed  for 12
                     hours, and for 7 days, then dissected and frozen (Akhunov
                     in J  Dvorak Lab). Total RNA was prepared from sheath
                     tissue, equal quantities  of RNA were pooled from the two
                     samples, polyA was purified from the  pooled RNA, a cDNA
                     library was made, and the cDNA clones were in vivo
                     excised to give pBluescript phagemids in the TJ Close lab
                     at the  University of California, Riverside (Akhunov, Chin
                     , Choi, Close, Fenton,  Kianian, Otto, Simons, Zhang).
                     Plasmid DNA preparations and DNA  sequencing were
                     performed in the OD Anderson lab (all other authors)."
BASE COUNT       62 a     89 c     86 g     60 t      3 others
ORIGIN

  Query Match             31.5%;  Score 31.8;  DB 13;  Length 300;
  Best Local Similarity   56.4%;  Pred. No. 57;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           |||||  |   |   | | ||| |||||||||   | | ||| | |  |||| ||||
Db       6 GTCGACGCCGCGNTGGNGGTGGCTTCGTCGGTGCCGGCGCCGTTCCCGGGGGAGGTGGCA 65

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           ||||| | || |  |  | |    ||  | || | |  | |
Db      66 GCCGCCGCCGCCGACTCCTTCTTCCCCGTCGCCGCGCTGCA 106


RESULT 6
BI875065
LOCUS       BI875065                572 bp    mRNA    linear   EST 11-OCT-2001
DEFINITION  963121E12.y1 C. reinhardtii CC-1690, Stress condition I, normalized
            , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI875065
VERSION     BI875065.1  GI:16073069
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
```

```
          ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE   Analyses of the Chlamydomonas reinhardtii Genome: A Model,
          Unicellular System for Analyzing Gene Function and Regulation in
          Vascular Plants. Project: 963
  JOURNAL Unpublished (2001)
COMMENT   Contact: Charles Hauser
          DCMB Box 91000
          Duke University
          Durham, NC 27708-1000
          Tel: 919 613 8159
          Fax: 919 613 8177
          Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, Stress condition I,
                     normalized, Lambda Zap II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
                     1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
                     , 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
                     purified from each sample, pooled and cDNA synthesized.
                     The cDNA was directionally cloned into lambda Zap II
                     (Stratagene) in the EcoRI (5') and XhoRI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      115 a    159 c    189 g    109 t
ORIGIN

  Query Match           31.1%;  Score 31.4;  DB 13;  Length 572;
  Best Local Similarity   64.4%;  Pred. No. 81;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      24 TTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCG 83
           ||| || |  | || |  || || |  ||||||||| ||||| |   ||||||| ||
Db     124 TTCCTCCGCTGGGCTCTTGCAGCCACTGGTGGTGCAAGCCGCGGTGTTCAGCGTTATTGC 183

Qy      84 TCCACTGGCGGAG 96
           || |   || |||
Db     184 TCAAGATGCTGAG 196


RESULT 7
BI336276
LOCUS       BI336276                484 bp    mRNA    linear   EST 01-SEP-2001
DEFINITION  AR094C04SPBC04S Porcine Spleen cDNA library Sus scrofa cDNA, mRNA
            sequence.
ACCESSION   BI336276
VERSION     BI336276.1  GI:15417572
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Rink,A., Santschi,E.M. and Beattie,C.W.
  TITLE     Amplified, Normalized cDNA Libraries from a Porcine Model of
            Orthopedic Implant Associated Staphylococcus aureus Infection
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Rink A
            Department of Animal Biotechnology
            College of Agriculture, Biotechnology and Natural Resources,
            University of Nevada, Reno
            MS 202, FA 103, 1664 N Virginia St, Reno, NV 89557-0236, USA
            Tel: 775 784 1705
            Fax: 775 784 1375
            Email: arink@cabnr.unr.edu
            Tissues and cells are derived from a porcine model for
            implant-associated infection using 1000 cfu of Staphylococcus
            aureus in a tibial transection, reduced and internally fixed with a
            dynamic compression plate. NOTE: The sequences contain a 'cDNA
            adapter' etween the EcoRI site and the start of the EST. The
            adapter sequence is 'AATTCGGCACGAG'.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone_lib="Porcine Spleen cDNA library"
                     /tissue_type="Spleen"
                     /cell_type="mixed"
                     /dev_stage="control, 5 month old castrated male"
                     /lab_host="SOLR"
                     /note="Vector: pBSK; Site_1: Eco RI; Site_2: XhoI; Tissues
                     and cells are derived from a porcine model for
                     implant-associated infection using 1000 cfu of
                     Staphylococcus aureus in a tibial transection, reduced and
                     internally fixed with a dynamic compression plate. NOTE:
                     The sequences contain a 'cDNA adapter' between the EcoRI
                     site and the start of the EST. The adapter sequence is
                     'AATTCGGCACGAG'."
BASE COUNT       90 a    138 c    158 g     82 t     16 others
ORIGIN

  Query Match           30.7%;  Score 31;  DB 13;  Length 484;
  Best Local Similarity   59.8%;  Pred. No. 1e+02;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
           | |  ||   ||||| |||| | | | |||  | | ||| ||||||    ||||| |  ||
Db     137 TGGGCCCAGAGACAGAGCTGCTGTGGCCGGGGGCGGCCCTGCTGCTGCTGGTGGGGGCCG 196

Qy      62 CCGCTGGCGTCAGCGTCATGCGTCCAC 88
           | || ||| |  | |||    || | |
Db     197 CGGCCGGCCTGTGTGTCCGCTGTTCCC 223


RESULT 8
BF078123
LOCUS       BF078123                547 bp    mRNA    linear   EST 18-OCT-2000
DEFINITION  228383 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF078123
VERSION     BF078123.1  GI:10871953
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 547)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 49  row: J  column: 24
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
```

```
     source          1. .547
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      109 a    159 c    180 g     99 t
ORIGIN

  Query Match             30.7%;  Score 31;  DB 12;  Length 547;
  Best Local Similarity   59.8%;  Pred. No. 1e+02;
  Matches   52;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
           | |  ||   ||||| |||| | | | |||  | | ||| |||||    ||||| |  ||
Db     139 TGGGCCCAGAGACAGAGCTGCTGTGGCCGGGGGCGGCCCTGCTGCTGCTGGTGGGGGCCG 198

Qy      62 CCGCTGGCGTCAGCGTCATGCGTCCAC 88
           | || ||| |  | |||    || | |
Db     199 CGGCCGGCCTGTGTGTCCGCTGTTCCC 225


RESULT 9
AV630710/c
LOCUS       AV630710                413 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV630710 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LCL083a07_r 5', mRNA sequence.
ACCESSION   AV630710
VERSION     AV630710.1  GI:10793344
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 413)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .413
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCL083a07_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       75 a    135 c    152 g     51 t
ORIGIN

  Query Match             30.5%;  Score 30.8;  DB 10;  Length 413;
  Best Local Similarity   63.5%;  Pred. No. 1.1e+02;
  Matches   47;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      23 TTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGC 82
           ||| | |  | | ||| |||||||    | || |||   | || |||| | |  ||||||
Db     224 TTTTGCCCTTGGGGCCGCCGCTGCCGTTGCTGTTGCCGCCGGCAGGCGGCGGGTTCATGC 165

Qy      83 GTCCACTGGCGGAG 96
           |||  | | ||| |
Db     164 GTCGGCCGCCGGGG 151


RESULT 10
BQ294770
LOCUS       BQ294770                479 bp    mRNA    linear   EST 16-MAY-2002
DEFINITION  WHE2854_B10_D20ZS Wheat unstressed root tip cDNA library Triticum
            aestivum cDNA clone WHE2854_B10_D20, mRNA sequence.
ACCESSION   BQ294770
VERSION     BQ294770.1  GI:20809751
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 479)
  AUTHORS   Anderson,O.D., Chao,S., Chin,A., Close,T.J., Crossman,C., Gustafson
            ,P., Lazo,G.R., Pham,J., Rausch,C.J., Ross,K., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Unstressed root tip cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .479
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE2854_B10_D20"
                     /clone_lib="Wheat unstressed root tip cDNA library"
                     /tissue_type="Root tip at 1.0 to 1.5 mm stage"
                     /dev_stage="Four-day old seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown under
                     hydroponic conditions for four days. Root tips were
                     excised and snap frozen (Ross and Gustafson) and total RNA
                     was prepared at University of Missouri, Columbia. Poly(A)
                     RNA was purified, a cDNA library was made, and the cDNA
                     clones were in vivo excised to give pBluescript SK(-
                     )phagemids in the TJ Close lab (Chin and Close) at the
                     University of California, Riverside. Plasmid DNA
                     preparations and DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT       60 a    217 c    115 g     87 t
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 14;  Length 479;
  Best Local Similarity   56.4%;  Pred. No. 1.3e+02;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           |||||  |   |   ||| ||| |||||||||   | | ||| | |  |||| ||||
Db     321 GTCGACGCCGCGGAGGCGGTGGCTTCGTCGGTGCCGGCGCCGTTCCCGGGGGAGGTGGCA 380

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           ||||| |  | |  |  | |    ||  | || | |  | |
Db     381 GCCGCCGCTGCCGACTCCTTTTTCCCCGTCGCCGCGCTGCA 421


RESULT 11
AW566146/c
LOCUS       AW566146                523 bp    mRNA    linear   EST 10-MAR-2000
```

```
DEFINITION  660062F05.y1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW566146
VERSION     AW566146.1  GI:7227505
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660062  row: F  column: 05.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
BASE COUNT      105 a    156 c    131 g    130 t      1 others
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 10;  Length 523;
  Best Local Similarity   58.1%;  Pred. No. 1.3e+02;
  Matches   54;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       7 CCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
           |||||||| |||    |   |  |||  ||||| | | || | |  |  |    |||| |
Db     111 CCGATGACGGCGGGTGCGACAGCGGAGGAGCCGCTGGTGGGGGACGGAGCCGGCGCCTCC 52

Qy      67 GGCGTCAGCGTCATGCGTCCACTGGCGGAGACG 99
           |  | |||||  || |    |   ||||| ||||
Db      51 GCGGCCAGCGCCACGGTGGCCGAGGCGGCGACG 19


RESULT 12
BH782972/c
LOCUS       BH782972                590 bp    DNA     linear   GSS 28-MAR-2002
DEFINITION  fzmb011f017h01k0 fzmb filtered library Zea mays genomic clone
            fzmb011f017h01 5', DNA sequence.
ACCESSION   BH782972
VERSION     BH782972.1  GI:19786407
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Budiman,M.A., Freese,R.G., Bedell,J.A., Nunberg,A.N. and Lakey,N.D.
  TITLE     GeneThresher methylation filtered genomic sequences from maize
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Bedell JA
            Orion Genomics, LLC
            4041 Forest Park Ave, St. Louis, MO 63108, USA
            Tel: 314 615 6979
            Fax: 314 615 5975
            Email: jbedell@oriongenomics.com
            Plate: fzmb011f017  row: h  column: 01
            Seq primer: SK reverse
            Class: shotgun
            High quality sequence stop: 590.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Zea mays"
                     /cultivar="MO17"
                     /db_xref="taxon:4577"
                     /clone="fzmb011f017h01"
                     /clone_lib="fzmb filtered library"
                     /note="Organ: leaf; Vector: pBCSK(-); Site_1: HincII; DNA
                     prepared from purified nuclei was randomly sheared,
                     end-repaired,  size fractionated to enrich for the 0.5 to
                     5 kb fraction, ligated into HincII-digested pBCSK(-)
                     vector and electroporated into E. coli cells."
BASE COUNT       98 a    210 c    176 g    106 t
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 17;  Length 590;
  Best Local Similarity   56.4%;  Pred. No. 1.3e+02;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           ||||   ||       ||| || |||  ||| |||||  || |  ||   ||| || ||||
Db     242 GTCGCTACGTGCCGCACGCCGGGTTCCGCGGCCGAGCTCCAGAAGCTCAGGGAGGAGCAC 183

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           ||  | ||    |  |  | |||  || || |  |  || ||
Db     182 GCGCCCGGGCGCCTCACCGTGCTGCCGCTCGACGTCACCGA 142


RESULT 13
AI621801
LOCUS       AI621801                610 bp    mRNA    linear   EST 22-APR-1999
DEFINITION  486026B10.x1 486 - leaf primordia cDNA library from Hake lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI621801
VERSION     AI621801.1  GI:4646726
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 486026  row: B  column: 10.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="486 - leaf primordia cDNA library from Hake
                     lab"
                     /tissue_type="leaf primordia"
                     /dev_stage="P7-P11 leaf"
                     /lab_host="E.coli XL1-Blue MFR'"
                     /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA
                     library."
BASE COUNT       82 a    226 c    189 g    107 t      6 others
```

```
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 9;  Length 610;
  Best Local Similarity   65.2%;  Pred. No. 1.3e+02;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      12 GACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGT 71
           | | |||||| || |||| |   ||| ||||| | | |||| | | |||| |||   | |
Db       9 GGCGGCGCTGCTTACGTCCGGGCAGCTCCCGCGGAGTGGGGAGCTCCACGACGCGCTCAT 68

Qy      72 CAGCGTCAT 80
           | | | | |
Db      69 CCGGGCCCT 77


RESULT 14
BG755515
LOCUS       BG755515                879 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602713891F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4854142 5',
            mRNA sequence.
ACCESSION   BG755515
VERSION     BG755515.1  GI:14066168
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 879)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1701  row: e  column: 23
            High quality sequence stop: 707.
FEATURES             Location/Qualifiers
     source          1. .879
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4854142"
                     /clone_lib="NIH_MGC_48"
                     /tissue_type="primary B-cells from tonsils (cell line)"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      196 a    264 c    282 g    137 t
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 12;  Length 879;
  Best Local Similarity   62.3%;  Pred. No. 1.4e+02;
  Matches   48;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      25 TCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGT 84
           ||  | |||| |  ||| |  ||||||| |||| |||| || | || |  || |||  |
Db     271 TCCCCCGTCGTGATCCCCCCACGAGGGGAGGTGAACGCGGCCGACGCCTTCGACATTGGC 330

Qy      85 CCACTGGCGGAGACGGA 101
            |  | |  |||  |||
Db     331 TCCTTCGATGAGGAGGA 347


RESULT 15
AY107409
LOCUS       AY107409                1024 bp    mRNA    linear   HTC 25-MAY-2002
DEFINITION  Zea mays PCO129924 mRNA sequence.
ACCESSION   AY107409
VERSION     AY107409.1  GI:21210487
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1024)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1024)
  AUTHORS   Coe,E.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .1024
                     /organism="Zea mays"
                     /db_xref="MaizeDB:637566"
                     /db_xref="taxon:4577"
                     /clone="PCO129924"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      186 a    295 c    332 g    211 t
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 11;  Length 1024;
  Best Local Similarity   62.3%;  Pred. No. 1.5e+02;
  Matches   48;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           ||||  ||| | |    ||||||  || ||| |  | ||   ||||  ||||||||||
Db      81 GTCGCCCCGCTCAACAAGCTGGTGCCGACGGCCACGGCCAGCCTGCACGGGGTGGTGCAG 140

Qy      61 GCCGCTGGCGTCAGCGT 77
           || |   | | | || |
Db     141 GCGGACCGGGGCTGCCT 157


Search completed: November  5, 2002, 16:18:03
Job time : 1118.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400
Perfect score: 101
Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactggcggagacgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	29.6	29.3	2787	10	US-09-815-242-6320	Sequence 6320, Ap
c 2	28.6	28.3	258	10	US-09-923-876-949	Sequence 949, App
c 3	28	27.7	336	10	US-09-954-456-675	Sequence 675, App
c 4	28	27.7	336	10	US-09-880-107-1676	Sequence 1676, Ap
c 5	27.8	27.5	217	10	US-09-923-876-2040	Sequence 2040, Ap
c 6	27.8	27.5	229	10	US-09-923-876-3118	Sequence 3118, Ap
7	27.8	27.5	594	10	US-09-815-242-7685	Sequence 7685, Ap
c 8	27.4	27.1	254	10	US-09-923-876-2583	Sequence 2583, Ap
c 9	27.4	27.1	2084	10	US-09-822-849A-526	Sequence 526, App
10	27.4	27.1	2742	10	US-09-815-242-4163	Sequence 4163, Ap
c 11	27.2	26.9	216	10	US-09-923-876-377	Sequence 377, App
c 12	27.2	26.9	257	10	US-09-923-876-336	Sequence 336, App
c 13	27.2	26.9	265	10	US-09-923-876-3689	Sequence 3689, Ap
c 14	27.2	26.9	274	10	US-09-923-876-3703	Sequence 3703, Ap
15	27.2	26.9	361	10	US-09-864-761-27835	Sequence 27835, A
16	27.2	26.9	489	10	US-09-864-761-11210	Sequence 11210, A
17	27.2	26.9	1185	10	US-09-974-300-911	Sequence 911, App
c 18	27.2	26.9	7373	10	US-09-764-853-896	Sequence 896, App
c 19	27	26.7	264	10	US-09-923-876-698	Sequence 698, App
20	27	26.7	348	10	US-09-864-761-31578	Sequence 31578, A
21	27	26.7	390	10	US-09-864-761-15047	Sequence 15047, A
22	26.8	26.5	4965	10	US-09-816-087-3	Sequence 3, Appli
23	26.8	26.5	12425	12	US-10-023-529-50	Sequence 50, Appl
24	26.8	26.5	12425	12	US-10-023-523-50	Sequence 50, Appl
c 25	26.4	26.1	243	10	US-09-923-876-2285	Sequence 2285, Ap
26	26.4	26.1	822	10	US-09-893-737-227	Sequence 227, App
27	26.4	26.1	2366	10	US-09-758-575-1	Sequence 1, Appli
28	26.4	26.1	2366	10	US-09-758-575-8	Sequence 8, Appli
c 29	26.4	26.1	5403	10	US-09-745-008-33	Sequence 33, Appl
c 30	26.2	25.9	224	10	US-09-923-876-2448	Sequence 2448, Ap
c 31	26.2	25.9	488	10	US-09-864-761-5557	Sequence 5557, Ap
c 32	26.2	25.9	495	12	US-10-033-528-1835	Sequence 1835, Ap
33	26.2	25.9	2922	10	US-09-925-302-73	Sequence 73, Appl
34	26.2	25.9	2994	12	US-10-044-090-356	Sequence 356, App
35	26	25.7	266	10	US-09-923-876-4826	Sequence 4826, Ap
36	26	25.7	529	10	US-09-764-877-975	Sequence 975, App
c 37	26	25.7	585	10	US-09-870-162A-10	Sequence 10, Appl
38	26	25.7	5025	10	US-09-960-253-176	Sequence 176, App
c 39	26	25.7	14272	10	US-09-870-162A-23	Sequence 23, Appl
40	25.8	25.5	1203	10	US-09-812-272-1	Sequence 1, Appli
41	25.8	25.5	1203	12	US-10-037-616-9	Sequence 9, Appli
c 42	25.8	25.5	1277	10	US-09-880-107-2230	Sequence 2230, Ap
c 43	25.6	25.3	242	10	US-09-923-876-2475	Sequence 2475, Ap
c 44	25.6	25.3	244	10	US-09-923-876-2527	Sequence 2527, Ap
c 45	25.6	25.3	248	10	US-09-923-876-1166	Sequence 1166, Ap

ALIGNMENTS

RESULT 1
US-09-815-242-6320
; Sequence 6320, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6320
; LENGTH: 2787
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2787)
US-09-815-242-6320

```
  Query Match             29.3%;  Score 29.6;  DB 10;  Length 2787;
  Best Local Similarity   57.6%;  Pred. No. 2.9;
  Matches   53;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
           |||||||  ||   ||| |||||    || | |  |  |||||||  | |||      ||
Db     275 TCGAACCCTTGCACGCGATGGTTAAAGCGATGGGACTGCCGCTGCTGGCGGTTTCTGGCG 334

Qy      62 CCGCTGGCGTCAGCGTCATGCGTCCACTG3CG 93
             |  |  | |  ||| ||  || | ||||||
Db     335 TAGAAGCGGACGACGTTATCGGTACTCTG3CG 366


RESULT 2
US-09-923-876-949/c
; Sequence 949, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/92?,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 949
;   LENGTH: 258
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700157856H1
;   NAME/KEY: unsure
;   LOCATION: 17, 38, 52, 75, 85, 92, 101, 115, 175, 196, 239
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-949

  Query Match             28.3%;  Score 28.6;  DB 10;  Length 258;
  Best Local Similarity   54.2%;  Pred. No. 5.1;
  Matches   52;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     |||||    ||| |
Db     166 AGCCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGNCGGCGCAG 107

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGG?GACGG 100
           |  |   ||  ||| ||   |  |||||||  |||
Db     106 CGAGGNCCATGGTCNTGATGCNGCTGGCGGCNGCGG 71


RESULT 3
US-09-954-456-675/c
; Sequence 675, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,?56
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,05?
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,92?
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 675
;   LENGTH: 336
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: n=a,t,g or c
US-09-954-456-675

  Query Match             27.7%;  Score 28;  DB 10;  Length 336;
  Best Local Similarity   64.5%;  Pred. No. 7.5;
  Matches   40;  Conservative    1;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           ||||||  ||  |:|| |  | ||| ||| ||||| || || | || | | |  | |||
Db     301 GAGCCCAGGCCCCRAGTGAAGCTGCTCGCAGCTGGAGTGAGAGGCAGGTGGCGCCGGGCC 242

Qy      94 GA 95
           ||
Db     241 GA 240


RESULT 4
US-09-880-107-1676/c
; Sequence 1676, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1676
;   LENGTH: 336
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 D51133
;   NAME/KEY: unsure
;   LOCATION: (1)..(336)
;   OTHER INFORMATION: n = a or c or g or t
US-09-880-107-1676

  Query Match             27.7%;  Score 28;  DB 10;  Length 336;
  Best Local Similarity   64.5%;  Pred. No. 7.5;
  Matches   40;  Conservative    1;  Mismatches   21;  Indels    0;  Gaps    0;
```

```
Qy      34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           ||||||  ||  |:|| |  | ||| ||| ||||| || || | || | | |  | |||
Db     301 GAGCCCAGGCCCCRAGTGAAGCTGCTCGCAGCTGGAGTGAGAGGCAGGTGGCGCCGGGCC 242

Qy      94 GA 95
           ||
Db     241 GA 240


RESULT 5
US-09-923-876-2040/c
; Sequence 2040, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 2040
;   LENGTH: 217
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700159955H1
;   NAME/KEY: unsure
;   LOCATION: 206
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-2040

  Query Match             27.5%;  Score 27.8;  DB 10;  Length 217;
  Best Local Similarity   55.2%;  Pred. No. 8.4;
  Matches   53;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     ||||||  ||| |
Db     216 AGCCGCCGCCNTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCGG 157

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |  |   |   |||  ||   |  |||||||  | |||
Db     156 CGAGGACCGTGGTCTTGATGCTGCTGGCGGCGGCGG 121


RESULT 6
US-09-923-876-3118/c
; Sequence 3118, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 3118
;   LENGTH: 229
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700161689H1
;   NAME/KEY: unsure
;   LOCATION: 37, 97, 177, 211, 216, 226
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-3118

  Query Match             27.5%;  Score 27.8;  DB 10;  Length 229;
  Best Local Similarity   55.2%;  Pred. No. 8.5;
  Matches   53;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     ||||||  ||| |
Db     155 AGCCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCNG 96

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |  |   |   |||  ||   |  |||||||  | |||
Db      95 CGAGGACCGTGGTCTTGATGCTGCTGGCGGCGGCGG 60


RESULT 7
US-09-815-242-7685
; Sequence 7685, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7685
;   LENGTH: 594
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(594)
US-09-815-242-7685

  Query Match             27.5%;  Score 27.8;  DB 10;  Length 594;
  Best Local Similarity   55.8%;  Pred. No. 8.7;
  Matches   53;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           || ||||||  |  | |     |||| |  ||||||  | |  | |    ||      ||
Db     115 GTGGAACCGGAGGAAACCGGTCTTTCCTTCGTCGAGAACGCCATCCTCAAGGCACGCAAC 174
```

```
Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGA 95
           |||||  || ||  || | |||   | ||||| ||
Db     175 GCCGCACGCCTCTCCGGCCTGCCGGCGCTGGCCGA 209


RESULT 8
US-09-923-876-2583/c
; Sequence 2583, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 2583
;   LENGTH: 254
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700160844H1
;   NAME/KEY: unsure
;   LOCATION: 173
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-2583

  Query Match             27.1%;  Score 27.4;  DB 10;  Length 254;
  Best Local Similarity   55.3%;  Pred. No. 11;
  Matches   52;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       7 CCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
           |||  | |  || |||   | |||     |||  |||||     ||||||   ||| ||
Db     254 CCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCAGCG 195

Qy      67 GGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |   ||  ||| ||   |   |||||| | |||
Db     194 AGGACCATGGTCTTGATGCTGNTGGCGGCGGCGG 161


RESULT 9
US-09-822-849A-526/c
; Sequence 526, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
;  APPLICANT: Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  APPLICANT:  Genetics Institute, Inc.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6403
;  CURRENT APPLICATION NUMBER: US/09/822,849A
;  CURRENT FILING DATE:  2001-09-04
;  PRIOR APPLICATION NUMBER: 60/195,582
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 598
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 526
;   LENGTH: 2084
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-849A-526

  Query Match             27.1%;  Score 27.4;  DB 10;  Length 2084;
  Best Local Similarity   57.6%;  Pred. No. 12;
  Matches   49;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       7 CCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
           | ||| |  | ||  | | |  ||||| ||  ||||| | | | || |||  || | | |
Db    1876 CTGATAATTGGGCACGGTGCCCCGGTCCAGGTCCCGCAGAGTGAGGGGGTCAACCCGGGT 1817

Qy      67 GGCGTCAGCGTCATGCGTCCACTGG 91
           |  |       || ||| ||| || |
Db    1816 GCAGAACCATTCCTGCCTCCGCTTG 1792


RESULT 10
US-09-815-242-4163
; Sequence 4163, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4163
;   LENGTH: 2742
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4163

  Query Match             27.1%;  Score 27.4;  DB 10;  Length 2742;
  Best Local Similarity   55.9%;  Pred. No. 12;
  Matches   52;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           ||||| ||| ||   ||    ||      |  |||  |  |||||||   | |||| |
Db     265 GTCGAGCCGCTGCATGCCAGCGTCCGCGCCCTCGGCCTGCCGCTGCTGTGCGTGGAGGGG 324

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           | ||  | || |  ||| ||  |  | ||||||
Db     325 GTCGAGGCCGACGACGTGATCGGCACCCTGGCG 357


RESULT 11
US-09-923-876-377/c
; Sequence 377, Application US/09923876
; Patent No. US20020013958A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 377
;   LENGTH: 216
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700156836H1
US-09-923-876-377

  Query Match             26.9%;  Score 27.2;  DB 10;  Length 216;
  Best Local Similarity   55.2%;  Pred. No. 12;
  Matches   53;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     ||||||   ||| |
Db     216 AGCCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCGG 157

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |  |   |   ||| ||   |  ||||||| | |||
Db     156 CGAGGACCGTGGTCTTGATGCTGCTGGCGGCGGCGG 121


RESULT 12
US-09-923-876-336/c
; Sequence 336, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 336
;   LENGTH: 257
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700156768H1
;   NAME/KEY: unsure
;   LOCATION: 215, 240, 244, 246
;   OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-336

  Query Match             26.9%;  Score 27.2;  DB 10;  Length 257;
  Best Local Similarity   56.2%;  Pred. No. 12;
  Matches   50;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      12 GACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGT 71
           | |  || |||   | |||     |||  |||||     ||||||   ||| ||  |
Db     236 GCCCTCGGTGGCCGCCTCGAANCCGCCGGCGCTGGCCTTGGTGGTCGGCGCAGCGAGGAC 177

Qy      72 CAGCGTCATGCGTCCACTGGCGGAGACGG 100
           ||  ||| ||   |  ||||||| | |||
Db     176 CATGGTCTTGATGCTGCTGGCGGCGGCGG 148


RESULT 13
US-09-923-876-3689/c
; Sequence 3689, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 3689
;   LENGTH: 265
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700405434H1
US-09-923-876-3689

  Query Match             26.9%;  Score 27.2;  DB 10;  Length 265;
  Best Local Similarity   55.2%;  Pred. No. 12;
  Matches   53;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     ||||||   ||| |
Db     216 AGCCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCGG 157

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |  |   |   ||| ||   |  ||||||| | |||
Db     156 CGAGGACCGTGGTCTTGATGCTGCTGGCGGCGGCGG 121


RESULT 14
US-09-923-876-3703/c
; Sequence 3703, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT:  Kamigaki, Laura Y. (Ito)
;  APPLICANT:  Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
;  FILE REFERENCE: PL-0012-1 CON
;  CURRENT APPLICATION NUMBER: US/09/923,876
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: 09/298,329
;  PRIOR FILING DATE: 1999-04-21
;  PRIOR APPLICATION NUMBER: 60/085,331
;  PRIOR FILING DATE: 1998-05-05
;  NUMBER OF SEQ ID NOS: 6332
;  SOFTWARE: PERL Program
; SEQ ID NO 3703
;   LENGTH: 274
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020013958A1 700405458H1
US-09-923-876-3703
```

```
  Query Match           26.9%;  Score 27.2;  DB 10;  Length 274;
  Best Local Similarity  55.2%;  Pred. No. 12;
  Matches   53;  Conservative    0;  Mismatches   43;  Indels    0;  Gaps    0;

Qy       5 AACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCG 64
           | |||  | |  || |||   | |||     |||  |||||     ||||||   ||| |
Db     217 AGCCGCCGCCCTCGGTGGCCGCCTCGAAGCCGCCGGCGCTGGCCTTGGTGGTCGGCGCGG 158

Qy      65 CTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
           |  |   |   ||| ||   |  ||||||| | |||
Db     157 CGAGGACCGTGGTCTTGATGCTGCTGGCGGCGG 122


RESULT 15
US-09-864-761-27835
; Sequence 27835, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00561
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00570
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,637
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27835
;   LENGTH: 361
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC003042.1
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.68
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.74
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.83
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71
;   OTHER INFORMATION: EST_HUMAN HIT: BE273103.1, EVALUE 1.00e-112
;   OTHER INFORMATION: SWISSPROT HIT: P39290, EVALUE 1.00e-06
US-09-864-761-27835

  Query Match           26.9%;  Score 27.2;  DB 10;  Length 361;
  Best Local Similarity  61.1%;  Pred. No. 13;
  Matches   44;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      30 GGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACT 89
           || |||||| |  ||||  ||||| ||   |   ||| || ||  || | |   ||||
Db      87 GGACGAGCCACAACTGCTGGGGGTCGTCGCCTGGGCTCGCCTCCCCGGCCTCTCTCCAAG 146

Qy      90 GGCGGAGACGGA 101
           | ||| | || |
Db     147 GCCGGGGCCGCA 158


Search completed: November  5, 2002, 23:00:16
Job time : 83.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400
Perfect score: 101
Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactggcggagacgga 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	29.6	29.3	33529	4	US-09-144-085-3	Sequence 3, Appli
	10	29	28.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
	11	29	28.7	44377	2	US-08-804-198-1	Sequence 1, Appli
	12	28.6	28.3	80161	3	US-09-036-987A-1	Sequence 1, Appli
	13	28.6	28.3	80161	4	US-09-370-700-1	Sequence 1, Appli
	14	28.4	28.1	13987	2	US-08-804-227C-13	Sequence 13, Appl
c	15	28	27.7	1485	4	US-09-372-422A-39	Sequence 39, Appl
c	16	28	27.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
c	17	28	27.7	28958	1	US-08-456-837-6	Sequence 6, Appli
c	18	28	27.7	28958	1	US-08-457-342-6	Sequence 6, Appli
c	19	28	27.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
c	20	28	27.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
c	21	28	27.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
c	22	28	27.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
c	23	28	27.7	28958	1	US-08-729-214-6	Sequence 6, Appli
c	24	28	27.7	28958	3	US-09-028-934-6	Sequence 6, Appli
c	25	28	27.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	26	28	27.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	27	26.8	26.5	1332	1	US-08-660-765A-3	Sequence 3, Appli
	28	26.4	26.1	50937	4	US-09-428-517-1	Sequence 1, Appli
	29	26.2	25.9	667	3	US-09-010-809-13	Sequence 13, Appl
c	30	26.2	25.9	2770	3	US-08-586-165-6	Sequence 6, Appli
c	31	26.2	25.9	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	32	26	25.7	585	4	US-09-516-914-10	Sequence 10, Appl
c	33	26	25.7	14272	4	US-09-516-914-23	Sequence 23, Appl
	34	26	25.7	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	35	25.8	25.5	1434	4	US-09-434-288-3	Sequence 3, Appli
	36	25.8	25.5	20235	1	US-07-642-734C-3	Sequence 3, Appli
	37	25.8	25.5	20235	3	US-08-439-009A-3	Sequence 3, Appli
	38	25.6	25.3	8438	1	US-07-945-283-1	Sequence 1, Appli
	39	25.4	25.1	532	4	US-09-449-218D-15	Sequence 15, Appl
	40	25.4	25.1	10095	3	US-08-822-586-45	Sequence 45, Appl
	41	25.4	25.1	49272	1	US-08-614-770A-1	Sequence 1, Appli
	42	25.2	25.0	1851	1	US-08-455-543A-10	Sequence 10, Appl
	43	25.2	25.0	1851	2	US-08-193-078B-10	Sequence 10, Appl
	44	25.2	25.0	1851	2	US-08-223-305C-10	Sequence 10, Appl
	45	25.2	25.0	1851	2	US-08-149-097D-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 3.1e-19;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-19;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24300 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 24359

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  24360 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 24400


RESULT 9
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match              29.3%;  Score 29.6;  DB 4;  Length 33529;
  Best Local Similarity    56.0%;  Pred. No. 6.7;
  Matches   56;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
          |||   || ||   |||  | |   |  ||  ||  | ||||||  | | |||||  ||||
```

```
Db  24388 TCGCGGCGCTGCTCGCGGAGCTGGAGCAGGATGAAGCGCCGCTGAGGGCGGTGGCGCACC 24447

Qy     62 CCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          ||| ||| |  ||  |     ||| |||   |||   ||
Db  24448 TCGCGGGCATAGGCCGCCGTGTTCCGCTGCGAGAGCTCGA 24487


RESULT 10
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             28.7%;  Score 29;  DB 2;  Length 44377;
  Best Local Similarity   55.4%;  Pred. No. 9.9;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          | ||   || ||   | || | | ||  ||| ||| ||| |||   | |||| |  | |
Db  35820 GGCGGCGCGCTGCTGGAGCGGCTGTCCGCGGACGAACCCGCGCCCGGCGGGGCGCCGGAG 35879

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||| || |  ||   |  |   |||| | | |   ||||
Db  35880 CCCGCCGGGGGGAGGCCCGCGACCCCACCGCCCGCACCGGA 35920


RESULT 11
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             28.7%;  Score 29;  DB 2;  Length 44377;
  Best Local Similarity   55.4%;  Pred. No. 9.9;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          | ||   || ||   | || | | ||  ||| ||| ||| |||   | |||| |  | |
Db  35820 GGCGGCGCGCTGCTGGAGCGGCTGTCCGCGGACGAACCCGCGCCCGGCGGGGCGCCGGAG 35879

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||| || |  ||   |  |   |||| | | |   ||||
Db  35880 CCCGCCGGGGGGAGGCCCGCGACCCCACCGCCCGCACCGGA 35920
```

```
RESULT 12
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             28.3%;  Score 28.6;  DB 3;  Length 80161;
  Best Local Similarity   64.2%;  Pred. No. 13;
  Matches   43;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           |||   ||||||   | |||||| |||||||| || ||| |||    | |     | | |
Db  43852 GAGTGTCCGCTGACTGCGGTGGTACACGCCGCCGGGGTCGGCGAGGCGGGCGACGTAGTG 43911

Qy      94 GAGACGG 100
           |||| ||
Db  43912 GAGATGG 43918


RESULT 13
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             28.3%;  Score 28.6;  DB 4;  Length 80161;
  Best Local Similarity   64.2%;  Pred. No. 13;
  Matches   43;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           |||   ||||||   | |||||| |||||||| || ||| |||    | |     | | |
Db  43852 GAGTGTCCGCTGACTGCGGTGGTACACGCCGCCGGGGTCGGCGAGGCGGGCGACGTAGTG 43911

Qy      94 GAGACGG 100
           |||| ||
Db  43912 GAGATGG 43918


RESULT 14
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..13987
US-08-804-227C-13
```

```
  Query Match             28.1%;  Score 28.4;  DB 2;  Length 13987;
  Best Local Similarity   56.4%;  Pred  No. 13;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           | || || || | | |      |    | ||   || | |||||||   || |||||||||
Db   12788 GCCGTACGGAAGGCCGTCGCCGCGGTGCCGTCGGACCACCCGCTGACCGGTGTGGTGCAC 12847

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGG 94
            |||| |||||    | |    |  | ||| | |
Db   12848 ACCGCCGGCGTGCTGGACGACGGCGCCCTGACCG 12881


RESULT 15
US-09-372-422A-39/c
; Sequence 39, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
;  APPLICANT: Rudolf Jung
;  APPLICANT:  Francois Barrieu
;  TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
;  FILE REFERENCE: 0919
;  CURRENT APPLICATION NUMBER: US/09/372,422A
;  CURRENT FILING DATE:  1999-08-11
;  PRIOR APPLICATION NUMBER: US 60/098,692
;  PRIOR FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 39
;   LENGTH: 1485
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (133)...(1015)
US-09-372-422A-39

  Query Match             27.7%;  Score 28;  DB 4;  Length 1485;
  Best Local Similarity   60.5%;  Pred. No. 15;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       3 CGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGC 62
           ||  |||||||| | || |     || | || ||   ||  ||     | || ||  | |
Db     882 CGTGCCGATGACGGGGCCGAGGAAGTAGATCCAGAGGCCCGTGAAGACGTTGCTGGCCAC 823

Qy      63 CGCTGGCGTCAGCGTC 78
           ||| |||| |||||||
Db     822 CGCCGGCGCCAGCGTC 807


Search completed: November  5, 2002, 15:09:46
Job time : 214.737 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400
Perfect score: 101
Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactggcggagacgga 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	101	100.0	71989	21	AAA29349	Sorangium cellulos
	3	31.6	31.3	27541	22	AAD17185	Streptomyces nours
	4	31.6	31.3	125401	22	AAD17186	Streptomyces nours
	5	29.8	29.5	53789	19	AAV21187	Amycolatopsis medi
	6	29.6	29.3	2787	23	AAS52683	E. coli DNA for ce
	7	29.6	29.3	4851	22	AAH52062	Mycobacterium tube
	8	29.6	29.3	33529	23	AAS17367	DNA sequence of S.
	9	29	28.7	44377	18	AAT78508	Platenolide syntha
	10	29	28.7	44377	18	AAT80414	Platenolide syntha
	11	28.8	28.5	3236	22	AAK51552	Human polynucleoti
	12	28.8	28.5	9558	16	AAQ88228	Valencia orange ri
	13	28.6	28.3	9513	22	AAF88337	S. spinosa DNA fra
	14	28.6	28.3	50000	22	AAF88313	S. spinosa DNA fra
	15	28.6	28.3	50000	22	AAF88316	S. spinosa DNA fra
	16	28.6	28.3	80161	20	AAZ21501	DNA fragment of Sa
	17	28.4	28.1	13987	18	AAT80415	Hybrid srmG/tylG O
	18	28.2	27.9	109519	22	AAS08693	Micromonospora DNA
c	19	28	27.7	336	24	ABN95178	Gene #1676 used to
c	20	28	27.7	336	24	ABL65365	Lung cancer relate
	21	28	27.7	1515	24	ABQ91596	M. capsulatus gene
c	22	28	27.7	28598	17	AAT06769	Sorangium cellulos
c	23	28	27.7	28958	18	AAT89956	Sorangium cellulos
c	24	28	27.7	28958	21	AAA75299	DNA sequence of So
c	25	28	27.7	49377	19	AAV05287	The soraphen biosy
	26	28	27.7	88421	24	AAL40781	88421nt genomic DN
c	27	28	27.7	4411529	22	AAI99682	Mycobacterium tube
	28	27.8	27.5	594	23	AAS54048	Pseudomonas aerugi
c	29	27.8	27.5	1215	22	AAH25116	Nucleotide sequenc
c	30	27.8	27.5	2306	22	AAH25117	Genomic DNA encodi
c	31	27.8	27.5	109519	22	AAS08693	Micromonospora DNA
c	32	27.6	27.3	2532	24	ABQ90377	M. capsulatus gene
c	33	27.6	27.3	2532	24	ABQ72983	Methylococcus caps
c	34	27.6	27.3	3163	22	AAK52536	Human polynucleoti
	35	27.6	27.3	65140	22	AAD17184	Streptomyces nours
c	36	27.4	27.1	1371	23	AAI99933	Human alpha-2CAR v
c	37	27.4	27.1	1382	15	AAQ64890	Human derived adre
c	38	27.4	27.1	1383	23	AAI99931	Human alpha-2CAR e
	39	27.4	27.1	1869	22	AAI59876	Human polynucleoti
c	40	27.4	27.1	1978	22	AAI58090	Human polynucleoti
c	41	27.4	27.1	2084	24	ABK35388	Human cDNA encodin
	42	27.4	27.1	2742	23	AAS51581	Pseudomonas aerugi
c	43	27.4	27.1	3126	22	AAK94778	Human full-length
c	44	27.4	27.1	3948	24	AAS98203	DNA encoding plako
c	45	27.4	27.1	3949	22	AAS05388	Human caspase recr

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;    99WO-EP04171.
XX
PR   18-JUN-1998;    98US-0099504.
PR   24-SEP-1998;    98US-0101631.
PR   05-FEB-1999;    99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-18;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24300 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 24359

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   24360 GCCGCTGGCGTCAGCGTCATGCGTCCACTGCGGAGACGGA 24400


RESULT 3
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             31.3%;  Score 31.6;  DB 22;  Length 27541;
  Best Local Similarity   65.7%;  Pred. No. 13;
  Matches   46;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      26 CGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTC 85
           ||  | | || |  |||||||| | |||||||  |||||  | || ||||| |  |
Db   13476 CGCTGTTGGACCGGCCGCTGCGCGAGGTGGTGTTCGCCGAGGACGGCAGCGACGAGGCCG 13535

Qy      86 CACTGGCGGA 95
           |||||  |||
Db   13536 CACTGCTGGA 13545


RESULT 4
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
```

```
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match            31.3%;  Score 31.6;  DB 22;  Length 125401;
  Best Local Similarity   65.7%;  Pred. No. 14;
  Matches   46;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      26 CGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTC 85
           ||  | | || |  |||||||| | |||||||  |||||  | || ||||| |  |
Db   13476 CGCTGTTGGACCGGCCGCTGCGCGAGGTGGTGTTCGCCGAGGACGGCAGCGACGAGGCCG 13535

Qy      86 CACTGGCGGA 95
           |||||  |||
Db   13536 CACTGCTGGA 13545


RESULT 5
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
```

DR P-PSDB; AAW52845-W52850.
XX
PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
PS Claim 4; Page 53-102; 205pp; English.
XX
CC The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster DNA fragment from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
XX
SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

Query Match 29.5%; Score 29.8; DB 19; Length 53789;
Best Local Similarity 63.0%; Pred. No. 41;
Matches 46; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 16 GCGCTGGTTTCGTCGGTCGAGCCCCCGCTCCGAGGGGTGGTGCACGCCGCTGGCGTCAGC 75
Db 35428 GCGCTGCTGGTGCCGATCGAGCTCGACCTCCGCGAGGTGCGGGCCGACGCGGCCGGCGGC 35487

Qy 76 GTCATGCGTCCAC 88
Db 35488 GGCACGGTGCCGC 35500

RESULT 6
AAS52683
ID AAS52683 standard; DNA; 2787 BP.
XX
AC AAS52683;
XX
DT 13-FEB-2002 (first entry)
XX
DE E. coli DNA for cellular proliferation protein #405.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU34824.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 6320; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2787 BP; 734 A; 679 C; 769 G; 605 T; 0 other;

Query Match 29.3%; Score 29.6; DB 23; Length 2787;
Best Local Similarity 57.6%; Pred. No. 42;
Matches 53; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
Db 275 TCGAACCCTTGCACGCGATGGTTAAAGCGATGGGACTGCCGCTGCTGGCGGTTTCTGGCG 334

Qy 62 CCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
Db 335 TAGAAGCGGACGACGTTATCGGTACTCTGGCG 366

RESULT 7
AAH52062
ID AAH52062 standard; DNA; 4851 BP.
XX
AC AAH52062;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 116.
XX
KW Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US31152.
XX
PR 12-NOV-1999; 99US-0165086.
PR 12-NOV-1999; 99US-0165124.
PR 01-FEB-2000; 2000US-0179531.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Eisenberg D, Rotstein SH, Marcotte EM;
XX

DR WPI; 2001-329193/34.
DR P-PSDB; AAG81211.
XX
PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -
XX
PS Disclosure; Page 131-133; 207pp; English.
XX
CC This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of
CC an organism.
XX
SQ Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

Query Match 29.3%; Score 29.6; DB 22; Length 4851;
Best Local Similarity 64.7%; Pred. No. 43;
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 9 GATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCCCTGG 68
Db 3894 GTTGTTTGCCCAGCTGTCGCGGGAGTACCCGCCGGTGCGCGGGGTGATTCATGCCGCCGG 3953

Qy 69 CGTCAGCG 76
Db 3954 CGTGCTCG 3961

RESULT 8
AAS17367
ID AAS17367 standard; DNA; 33529 BP.
XX
AC AAS17367;
XX
DT 12-MAR-2002 (first entry)
XX
DE DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW acyl transferase domain; dehydratase domain; ketoreductase domain;
KW acyl carrier protein domain; pKOS28-26; ds.
XX
OS Sorangium cellulosum.
XX
PN US6280999-B1.
XX
PD 28-AUG-2001.
XX
PF 31-AUG-1998; 98US-0144085.
XX
PR 22-JAN-1998; 98US-0010809.
XX
PA (KOSA-) KOSAN BIOSCIENCE.
XX
PI Gustafsson C, Betlach MC, Ashley G, Julien B, Ziermann R;
XX
DR WPI; 2001-606536/69.
XX
PT Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT polyketide open reading frame encoding modules with one or more domains
PT such as ketosynthase, acyl transferase and acyl carrier protein domains
PT -
XX
PS Claim 4; Fig 1; 72pp; English.
XX
CC The present invention relates to the isolation of novel Sorangium
CC cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC encoding them. The polyketide synthases include catalytic domains such
CC as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC ketoreductase domain and acyl carrier protein domain. A host cell
CC comprising a PKS ORF (open reading frame) which encodes one or more
CC more PKS domains is useful for producing polyketide synthases from which
CC polyketides can be produced. The host cells are useful for constructing
CC a library, where each individual colony of the library represents a
CC colony with the ability to produce a particular PKS synthase and
CC ultimately a particular polyketide. The polyketides produced by these
CC colonies can be used collectively in a panel to represent a library or
CC may be assessed individually for activity. Colonies in the library are
CC also induced to produce the relevant synthases and thus to produce the
CC relevant polyketides to obtain a library of candidate polyketides which
CC can be screened for binding to desired targets such as receptors,
CC signalling proteins, etc. The present sequence represents the DNA
CC sequence of cosmid pKOS28-26 which encodes one or more domains of
CC S. cellulosum PKS.
CC Note: The present sequence is said to encode the functional domains
CC of S. cellulosum PKS which correspond to domains or domain subsets of
CC the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

Query Match 29.3%; Score 29.6; DB 23; Length 33529;
Best Local Similarity 56.0%; Pred. No. 46;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 2 TCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACG 61
Db 24388 TCGCGGCGCTGCTCGCGGAGCTGGAGCAGGATGAAGCGCCGCTGAGGGCGGTGGCGCACC 24447

Qy 62 CCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
Db 24448 TCGCGGGCATAGGCCGCCGTGTTCCGCTGCGAGAGCTCGA 24487

RESULT 9
AAT78508
ID AAT78508 standard; DNA; 44377 BP.
XX
AC AAT78508;
XX
DT 26-FEB-1998 (first entry)
XX
DE Platenolide synthase gene cluster.
XX
KW Platenolide synthase gene cluster; platenolide production; srmG gene;
KW multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS Streptomyces ambofaciens.
XX
FH Key Location/Qualifiers
FT CDS 350..14002
FT /*tag= a
FT /transl_except= (pos:350..352, aa:Met)
FT /note= "ORF1 encodes protein shown in AAW23716"
FT CDS 14046..20036
FT /*tag= b
FT /note= "ORF2 encodes protein shown in AAW23717"
FT CDS 20110..31284
FT /*tag= c
FT /transl_except= (pos:20111..20113, aa:Met)
FT /note= "ORF3 encodes protein shown in AAW23718"
FT CDS 31329..36071
FT /*tag= d
FT /note= "ORF4 encodes protein shown in AAW23719"
FT CDS 36155..41830
FT /*tag= e

```
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             28.7%;  Score 29;  DB 18;  Length 44377;
  Best Local Similarity   55.4%;  Pred. No. 67;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           | ||   || ||   | || | | ||  ||| ||| ||| |||   | |||| |  | |
Db   35820 GGCGGCGCGCTGCTGGAGCGGCTGTCCGCGGACGAACCCGCGCCCGGCGGGGCGCCGGAG 35879

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
            |||| || |  ||   |  |   |||| | | |   ||||
Db   35880 CCCGCCGGGGGGAGGCCCGCGACCCCACCGCCCGCACCGGA 35920


RESULT 10
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             28.7%;  Score 29;  DB 18;  Length 44377;
  Best Local Similarity   55.4%;  Pred. No. 67;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           | ||   || ||   | || | | ||  ||| ||| ||| |||   | |||| |  | |
Db   35820 GGCGGCGCGCTGCTGGAGCGGCTGTCCGCGGACGAACCCGCGCCCGGCGGGGCGCCGGAG 35879

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
            |||| || |  ||   |  |   |||| | | |   ||||
Db   35880 CCCGCCGGGGGGAGGCCCGCGACCCCACCGCCCGCACCGGA 35920


RESULT 11
AAK51552
ID   AAK51552 standard; cDNA; 3236 BP.
XX
AC   AAK51552;
XX
```

```
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 97.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US04098.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   20-JUN-2000; 2000US-0598075.
PR   19-JUL-2000; 2000US-0620325.
PR   01-SEP-2000; 2000US-0654936.
PR   15-SEP-2000; 2000US-0663561.
PR   20-OCT-2000; 2000US-0693325.
PR   30-NOV-2000; 2000US-0728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y,  Ma Y;
PI   Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM78419.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities,
PT   useful in diagnosis and gene therapy -
XX
PS   Claim 1; Page 717-720; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC   (AAM80020) are omitted as the relevant pages from the sequence listing
CC   were missing at the time of publication.
XX
SQ   Sequence 3236 BP; 656 A; 924 C; 1040 G; 613 T; 3 other;

  Query Match            28.5%;  Score 28.8;  DB 22;  Length 3236;
  Best Local Similarity   60.0%;  Pred. No. 68;
  Matches   48;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      17 CGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCG 76
           ||  | ||||||  || | |  | |||| || |  |  ||    || ||||  |  |||
Db      52 CGACGATTTCGTAAGTGGGGTGCACGCTTCGGGTTGGTGTCATGGCAGCTGCGGGGAGCC 111

Qy      77 TCATGCGTCCACTGGCGGAG 96
             || | |  |||||||||
Db     112 GCAAGAGGCGCCTGGCGGAG 131


RESULT 12
AAQ88228
ID   AAQ88228 standard; DNA; 9558 BP.
XX
AC   AAQ88228;
XX
DT   24-NOV-1995  (first entry)
XX
DE   Valencia orange ribosomal RNA gene.
XX
KW   Ribosomal RNA; rRNA; 18S; 28S; 5.8S; Valencia orange; RFLP analysis;
KW   DNA fingerprinting; restriction fragment length polymorphism; ds.
XX
OS   Citrus sp.
XX
FH   Key             Location/Qualifiers
FT   misc_RNA        1..1597
FT                   /*tag= a
FT                   /product= 18S_rRNA
FT   misc_RNA        1598..1848
FT                   /*tag= b
FT                   /function= spacer
FT   misc_RNA        1849..2011
FT                   /*tag= c
FT                   /product= 5.8S_rRNA
FT   misc_RNA        2012..2238
FT                   /*tag= d
FT                   /function= spacer
FT   misc_RNA        2239..5582
FT                   /*tag= e
FT                   /product= 28S_rRNA
FT   misc_RNA        5583..9396
FT                   /*tag= f
FT                   /function= spacer
FT   misc_RNA        7338..8291
FT                   /*tag= g
FT                   /function= probe
FT                   /note= "claimed"
FT   misc_RNA        9397..9558
FT                   /*tag= h
FT                   /product= 18S_rRNA
XX
PN   JP07059577-A.
XX
PD   07-MAR-1995.
XX
PF   23-AUG-1993;   93JP-0227806.
XX
PR   23-AUG-1993;   93JP-0227806.
XX
PA   (KIRI ) KIRIN BEVERAGE KK.
XX
DR   WPI; 1995-135901/18.
XX
PT   Valencia orange ribosomal RNA gene - also probes and primers
PT   derived from citrus fruit rRNA gene spacer regions, useful for
PT   RFLP analysis of citrus fruit
XX
PS   Claim 1; Page 7-10; 18pp; Japanese.
XX
CC   The ribosomal RNA gene has been isolated from Valencia oranges and
CC   sequenced (see AAQ88228). Primers based on sequences in the rRNA gene
CC   (see AAQ88229-Q88230 which are derived from nucleotides 1556-1575 and
CC   2257-2276 of AAQ88228, respectively) can be used to amplify regions of
CC   rRNA genes from various different types of citrus fruits. RFLP
CC   analysis and DNA fingerprinting of the amplified fragments allows
CC   different species to be distinguished.
XX
SQ   Sequence 9558 BP; 1853 A; 2675 C; 2996 G; 2034 T; 0 other;

  Query Match            28.5%;  Score 28.8;  DB 16;  Length 9558;
  Best Local Similarity   65.6%;  Pred. No. 72;
  Matches   42;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       9 GATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGG 68
             | ||   ||| ||||  ||   || | ||||  |||||| | | ||||||    ||||||
```

```
Db   7895 GCTGCGCGCGATGGTGCCGCGCGTGGCGCCATGCTGCGCGCGATGGTGCCGCGCGCTGG 7954

Qy     69 CGTC 72
          || |
Db   7955 CGCC 7958
```

RESULT 13
AAF88337
ID AAF88337 standard; DNA; 9513 BP.
XX
AC AAF88337;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF20, SEQ ID 45.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR P-PSDB; AAB70967.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7a; Page 209-226; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 20, encodes an S. spinosa polyketide synthase.
XX
SQ Sequence 9513 BP; 1392 A; 2759 C; 3637 G; 1725 T; 0 other;

```
  Query Match             28.3%;  Score 28.6;  DB 22;  Length 9513;
  Best Local Similarity   64.2%;  Pred. No. 81;
  Matches   43;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
          |||   ||||||   | ||||||  |||||||| || ||| |||    | |     | | |
Db   8434 GAGTGTCCGCTGACTGCGGTGGTACACGCCGCCGGGGTCGGCGAGGCGGGCGACGTAGTG 8493

Qy     94 GAGACGG 100
          |||| ||
Db   8494 GAGATGG 8500
```

RESULT 14
AAF88313
ID AAF88313 standard; DNA; 50000 BP.
XX
AC AAF88313;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 2.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7; Page 31-49; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine, trimethylrhamnose and polyketide synthase biosynthesis.
XX
SQ Sequence 50000 BP; 6867 A; 14165 C; 19274 G; 9694 T; 0 other;

```
  Query Match             28.3%;  Score 28.6;  DB 22;  Length 50000;
  Best Local Similarity   64.2%;  Pred. No. 87;
  Matches   43;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy     34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
          |||   ||||||   | ||||||  |||||||| || ||| |||    | |     | | |
Db  18481 GAGTGTCCGCTGACTGCGGTGGTACACGCCGCCGGGGTCGGCGAGGCGGGCGACGTAGTG 18540

Qy     94 GAGACGG 100
          |||| ||
Db  18541 GAGATGG 18547
```

RESULT 15
AAF88316
ID AAF88316 standard; DNA; 50000 BP.
XX
AC AAF88316;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 5.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7; Page 74-91; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine and trimethylrhamnose biosynthesis.
XX
SQ Sequence 50000 BP; 6944 A; 14312 C; 19170 G; 9574 T; 0 other;

Query Match 28.3%; Score 28.6; DB 22; Length 50000;
Best Local Similarity 64.2%; Pred. No. 87;
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```
Qy      34 GAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCTGGCGTCAGCGTCATGCGTCCACTGGCG 93
           |||   ||||||   | |||||| |||||||| || ||| |||    | |     | | |
Db   22857 GAGTGTCCGCTGACTGCGGTGGTACACGCCGCCGGGGTCGGCGAGGCGGGCGACGTAGTG 22916

Qy      94 GAGACGG 100
           |||| ||
Db   22917 GAGATGG 22923
```

Search completed: November 5, 2002, 13:08:37
Job time : 246.215 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_24300_24400
Perfect score: 101
Sequence: 1 gtcgaaccgatgacagcgct.....cgtccactggcggagacgga 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	58733	1	AF217189	AF217189 Sorangium
	2	101	100.0	68750	1	AF210843	AF210843 Sorangium
	3	101	100.0	68750	6	AR193029	AR193029 Sequence
	4	101	100.0	68750	6	AR199551	AR199551 Sequence
	5	101	100.0	68750	6	AR199559	AR199559 Sequence
	6	101	100.0	68750	6	AR199567	AR199567 Sequence
	7	101	100.0	68750	6	AR201097	AR201097 Sequence
	8	101	100.0	68750	6	AR208671	AR208671 Sequence
	9	101	100.0	71989	6	AR172664	AR172664 Sequence
	10	67.4	66.7	10910	1	AX024383	AX024383 Sequence
	11	67.4	66.7	10910	6	AX024276	AX024276 Sequence
	12	40.2	39.8	36905	1	AF509565	AF509565 Streptomy
	13	34.8	34.5	213050	1	AL646079	AL646079 Ralstonia
	14	34.4	34.1	6286	1	AF009336	AF009336 Streptomy
c	15	34.4	34.1	34713	1	SCF42	AL137165 Streptomy
c	16	34	33.7	45396	1	SCH35	AL078610 Streptomy
c	17	32.8	32.5	1671	1	AX024351	AX024351 Sequence
c	18	32.8	32.5	1671	6	AX024244	AX024244 Sequence
	19	32.8	32.5	2274	1	AX024349	AX024349 Sequence
	20	32.8	32.5	2274	6	AX024242	AX024242 Sequence
c	21	32.8	32.5	11992	1	AE005686	AE005686 Caulobact
	22	32.8	32.5	47713	1	AX024320	AX024320 Sequence
	23	32.8	32.5	47713	6	AX024213	AX024213 Sequence
c	24	32.6	32.3	2030	14	SHE012474	AJ012474 Cercopith
c	25	32.4	32.1	137785	2	AP005065	AP005065 Oryza sat
	26	32.4	32.1	176171	2	AP005067	AP005067 Oryza sat
	27	32.2	31.9	30277	1	AY054120	AY054120 Mycobacte
	28	32.2	31.9	104326	1	AB070940	AB070940 Streptomy
c	29	31.8	31.5	42816	1	SCC88	AL139298 Streptomy
	30	31.6	31.3	27541	6	AX211706	AX211706 Sequence
	31	31.6	31.3	123580	1	AF263912	AF263912 Streptomy
	32	31.6	31.3	125401	6	AX211739	AX211739 Sequence
c	33	31.4	31.1	150039	2	AC114018	AC114018 Rattus no
c	34	31.4	31.1	169733	2	AC068046	AC068046 Homo sapi
c	35	31	30.7	12902	1	AE011870	AE011870 Xanthomon
c	36	31	30.7	134155	2	AP005517	AP005517 Oryza sat
c	37	31	30.7	142278	2	OSJN00135	AL662942 Oryza sat
	38	31	30.7	148777	2	AP005410	AP005410 Oryza sat
	39	31	30.7	167617	2	AC091403	AC091403 Sus scrof
c	40	30.8	30.5	3059	14	AF078729	AF078729 Cervid he
	41	30.8	30.5	232941	2	AL670236	AL670236 Mus muscu
	42	30.8	30.5	254197	2	AF389853	AF389853 Mus muscu
c	43	30.6	30.3	257717	2	AC105653	AC105653 Rattus no
c	44	30.4	30.1	146893	2	CNS08C9A	AL732535 Oryza sat
	45	30.2	29.9	2000	1	AB015764	AB015764 Halobacte

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA
FEATURES Location/Qualifiers
source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"
CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRSERHLDPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLIPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFEPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"
gene 1998. .6263
/gene="epoA"
CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRPLERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPEDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
gene 6260. .10492
/gene="epoB"
CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene 10639. .16137
/gene="epoC"
CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 5e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24300 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGCTGCAC 24359

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||||
Db   24360 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 24400


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVCLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPAEVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIHIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251. .21749
                     /gene="epoB"
     CDS             16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match           100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
```

```
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  29912 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 29971

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  29972 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 30012


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match            100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24300 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 24359

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  24360 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 24400


RESULT 10
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match            66.7%;  Score 67.4;  DB 1;  Length 10910;
  Best Local Similarity  79.2%;  Pred. No. 3.5e-06;
  Matches   80;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          | ||| || ||||| |||||| |  |  |  ||||||||||| |||| ||||||||||||
Db   1350 GCCGATCCCATGACGGCGCTGCTGGCCGCCATCGAGCCCCCGTTGCGCGGGGTGGTGCAC 1409

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||  ||||||  |  | ||||||  ||||||||||||||
Db   1410 GCCGCCGGCGTCTTCCCCGTGCGTCACCTGGCGGAGACGGA 1450


RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match            66.7%;  Score 67.4;  DB 6;  Length 10910;
  Best Local Similarity  79.2%;  Pred. No. 3.5e-06;
  Matches   80;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
          | ||| || ||||| |||||| |  |  |  ||||||||||| |||| ||||||||||||
Db   1350 GCCGATCCCATGACGGCGCTGCTGGCCGCCATCGAGCCCCCGTTGCGCGGGGTGGTGCAC 1409

Qy     61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
          |||||  ||||||  |  | ||||||  ||||||||||||||
Db   1410 GCCGCCGGCGTCTTCCCCGTGCGTCACCTGGCGGAGACGGA 1450


RESULT 12
AF509565
LOCUS       AF509565               36905 bp    DNA     linear   BCT 06-JUN-2002
DEFINITION  Streptomyces sp. JP95 griseorhodin biosynthesis gene cluster,
            complete sequence.
ACCESSION   AF509565
```

```
VERSION     AF509565.1  GI:21039487
KEYWORDS    .
SOURCE      Streptomyces sp. JP95.
  ORGANISM  Streptomyces sp. JP95
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 36905)
  AUTHORS   Li,A. and Piel,J.
  TITLE     A gene cluster from a marine Streptomyces encoding the biosynthesis
            of the aromatic spiroketal polyketide griseorhodin A
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36905)
  AUTHORS   Li,A. and Piel,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Department of Bioorganics, Max-Planck
            Institute for Chemical Ecology, Winzerlaer Strasse 10, Beutenberg
            Campus, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1. .36905
                     /organism="Streptomyces sp. JP95"
                     /strain="JP95"
                     /isolate="gri"
                     /db_xref="taxon:195040"
                     /clone="6C2"
     gene            1245. .2066
                     /gene="grhR1"
     CDS             1245. .2066
                     /gene="grhR1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional activator GrhR1"
                     /protein_id="AAM33679.1"
                     /db_xref="GI:21039514"
                     /translation="MDIEILGSLLVSENGIPVMPTAPKVRQVLAILAFFAGQVVPIAV
                     LTDELWAGRPPRSARTTLQTYVLQLRELITVALPAPGRGKKALTAKDILVTCPGGYLL
                     NTRGGTSDVREFERLAGLGYRAVGDGDHAEGARLLGASLALWRGAPFADVQLGKQLEM
                     EAARLEESRLCALDQRIEADLRLGRHRELLAELTVLVSRYRTHENLCRQFMIALQRSG
                     RWSDALEVYRKLRGTLSRELGLEPSLALRQLQRSILLDHSQMTASDSVPRAVRAV"
     gene            2296. .3117
                     /gene="grhR2"
     CDS             2296. .3117
                     /gene="grhR2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcriptional activator GrhR2"
                     /protein_id="AAM33680.1"
                     /db_xref="GI:21039515"
                     /translation="MRIRVLGPLNADVNGVSVVPSAGKPRQVLSLLALYPGQIMPVPT
                     LMEEIWGAAPPVSALTTLQTYILQLRRRLGTAMGPEAPHTAKDVLATRHGGYVLQIPA
                     DSTDVHAYERKVREGQAAFEAGDDESAADTLREALALWQGPALVDVRCGPALSIEVLR
                     LEESRLVTVERRIDADLRLGRHCEISAELRDLTARHPHHEGLHSQAMVALYRSGRQAA
                     ALDLYRKLRVRLIEELGVEPSPQMQRMHQAMLSVDPALDLATGPRRGSTFDLYAA"
     gene            3315. .4130
                     /gene="grhD"
     CDS             3315. .4130
                     /gene="grhD"
                     /codon_start=1
                     /transl_table=11
                     /product="putative thioesterase GrhD"
                     /protein_id="AAM33656.1"
                     /db_xref="GI:21039491"
                     /translation="MAGTEPAVRLYCFAHAGAGDSVFHRWARSTGEGVEPVPVLLPGR
                     GGRRREPRLTDHDSLVAHLLDRVRVPAGQSYALYGHSLGAMVAHALAGALREEGLPLP
                     SFVAVGACPPPDAASRLSDACESSDEELVEALDRLGALPEGVAARGYWRRASLPVLRD
                     DLRLAGALRAAARASAPERGPLSVPLLAVSGSRDPLVSPATVGGWRRWTSGPVAARTV
                     PGDHFFVRGEALPRLIGRACRVVRRLEAGTGVGNGRWPAETVFPYIREPAVHP"
     gene            4323. .4583
                     /gene="grhC"
     CDS             4323. .4583
                     /gene="grhC"
                     /codon_start=1
                     /transl_table=11
                     /product="putative acyl carrier protein GrhC"
                     /protein_id="AAM33655.1"
                     /db_xref="GI:21039490"
                     /translation="MQQLDLPTLIGLLRTSGGSDTHDLNGDVLDVAFADLGYDSLCLL
                     QVTGVIERDAGVALDEQVLEEAETPRQYIAAVNRVLSPQATA"
     gene            4744. .5715
                     /gene="grhE"
     CDS             4744. .5715
                     /gene="grhE"
                     /codon_start=1
                     /transl_table=11
                     /product="putative cyclase GrhE"
                     /protein_id="AAM33657.1"
                     /db_xref="GI:21039492"
                     /translation="MPETAFHHTSHRVDIAAPAGVVYGLVADAVRWPLFLTPTVHVEA
                     LDLEGADQRLRIWALACDDVTSWVTRRTLDPAARRIDFRQEVPGPSEPGVTGSWVVEE
                     LPEGSSRLTLRSALPASATRTTPASWLERAEHDDLLALLADLKHLAERWATLDEHIVA
                     VEETLRVDGPAELAYDFLYRIENWPAQLPRLTRLDLVEETPGIQQVRMDLLADDGRTH
                     TARGVRVCFPHGGRIVYKLVEPPASTVAQAGEWSVEPDASGVTLRGRHQAVLRCPDAP
                     SDTEPRTESGGGAGPEVGRMLGRDSAVVLGLARRYAQSAVRVITPKP"
     gene            5777. .6580
                     /gene="grhF"
     CDS             5777. .6580
                     /gene="grhF"
                     /codon_start=1
                     /transl_table=11
                     /product="putative phosphopantetheinyl transferase GrhF"
                     /protein_id="AAM33658.1"
                     /db_xref="GI:21039493"
                     /translation="MPRLAPPVHVGGPDGPWNQLHRQWEDSGGVVVLATWGEWLPAVV
                     TDPRLRGLLGDDWPRYRRTADATVRHRFAAARFVLKHAAAAALETDPAELDLAYRPGG
                     RPYLRGVRGVDVSLAHTGELVVVGLSRTGRIGVDAEPLTRSLSVDVLLARACTPDERA
                     AVKTLGGIGGAAGVTAALLRLWTLKEAYTKALGCGLRLPFARFGVPWAVEGEAAAYGD
                     EGLLSPGGGPAGPRAWTFGAHRALGRYLVGVARHDGAPAAPLRPSSDRP"
     gene            6920. .8485
                     /gene="grhG"
     CDS             6920. .8485
                     /gene="grhG"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transcarboxylase component of acyl-CoA
                     carboxylase GrhG"
                     /protein_id="AAM33659.1"
                     /db_xref="GI:21039494"
                     /translation="MTTETAAASDPTAELAQLKQEAADGPDEQATARQHAKGKLTARE
                     RIDLLLDEGSFQEIEALRRHRASGFGLENKRPYTDGVVTGWGTVHGRTVFVYAHDFRI
                     FGGALGEAHAAKIHKIMDMALAAGAPLVSLNDGAGARIQEGVTALAGYGGIFRRNTLA
                     SGVIPQISVMLGPCAGGAAYSPALTDFIFMVRGTSQMFITGPDVVKAVTGEEITQNGL
                     GGADVHAAVSGVAAFAHDDEAACLEDVRFLLSLLPSNNRETAPHRAPDDPADRPTDAL
                     RDLVPANPNQAYDVRAVIEEIVDDGEFFEIQPDWASNIVCALTRLDGHVTGVVANQPA
                     AFAGVLDIHASEKAARFVQFCDAFSIPLVTLVDVPGFLPGVDQEHGGVIRHGAKLLYA
                     YCNATVPRVQLITRKAYGGAYIVMDSRSIGADVSLAWPTNEIAVMGAEGAANVIFRRE
                     IAAADDPDAVRAQRIKQYRTELMHPYYAAERGLVDDVIDPAATRSALIGALAMLRPKY
                     TDLPSRKHGNPPV"
     gene            8502. .8843
                     /gene="grhH"
     CDS             8502. .8843
                     /gene="grhH"
                     /codon_start=1
                     /transl_table=11
                     /product="GrhH"
                     /protein_id="AAM33660.1"
                     /db_xref="GI:21039495"
                     /translation="MTPTTTGAHTESAPNGTSGPHTEFGRNGTTEPAAGPGPAAVPGT
                     AISARPAWRISGGSPTAEETAALAAVLAVVLGRAAGGGRDAADAPGSTGTGWHRPPRV
                     RAPSPGTWRGR"
     gene            9055. .9423
                     /gene="grhI"
     CDS             9055. .9423
                     /gene="grhI"
                     /codon_start=1
                     /transl_table=11
                     /product="GrhI"
                     /protein_id="AAM33661.1"
```

```
                     /db_xref="GI:21039496"
                     /translation="MSDARLVGTWTTELEDDGKSVFGLASFTGDGGVTCYQVNAKNIG
                     LGNWESTGKDSFHYNFHILAVNPQGEHVGEAHVFVTGEFVSDDRWEGTGGANFYSPAG
                     DLLRGHEGSRVTARKFGIDK"
     gene            9482. .10885
                     /gene="grhO1"
     CDS             9482. .10885
                     /gene="grhO1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative FAD-binding oxidoreductase GrhO1"
                     /protein_id="AAM33667.1"
                     /db_xref="GI:21039502"
                     /translation="MLRDDLTGELVLPSDPGYRLAKQVQNTEYDAISPRAIAYCTSPD
                     DVLACVRHAREAGIPAHVRGAGHSFNGWSTGTGLVIDLSRMNHAAADGPTVSIGAGIQ
                     SLDALDALKRHDRQIITGTFPTVGQGGFLSGGGLGWQTRKFGVASDRIVSARVVLADG
                     RAVVASADEHPDLFWALRGGGGGNFGIVTAFELRPVDAPHLTAFETLWSFDDAADVLT
                     AWQSWAAGESDDLGTSLVVLPGMFGPGGRPVVKVWGVHHGAAEELERALDALTDLAGV
                     KPAHRETTATGPYADVMHEALCGSKTVQQCHRTGTGPEAVGHRHPYTRQAYRLTGRPV
                     TRAEAESLLRAWEPDLDHEHYLLCIALGGAANRVGRTETAYVHRDARFLTGYQIATRD
                     PEAVARGPVELDAWADRCAEALAPLATGSYINFPDSRTPPDWEADHYGENAARLRAVK
                     RAYDPDGFFRHGRSIGS"
     gene            10944. .11717
                     /gene="grhO2"
     CDS             10944. .11717
                     /gene="grhO2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative 3-oxoacyl-ACP reductase GrhO2"
                     /protein_id="AAM33669.1"
                     /db_xref="GI:21039504"
                     /translation="MKTGLTDKAVLVTGASRGIGRATALAYAAEGARVAITYHSDEAG
                     AKETAERVAAAGGTPHVVRYDLGDEDSVRSAVASVGAEWGGVDVLVAGAVEWGEAIPR
                     PGRKMPAFEEVPPAQWQRVLRTSVDGVFHTVQAVLPLMRDREWGRIVVLGAGLAETGL
                     PGASAYGAAKAALHGLVRSLAWEVGPAGILVNEVVPGQTLTENVLAHASPAFLENKAA
                     SLPSGRMNTPEDVARAIVFLGSAANGNIHGEALRVTGGL"
     gene            11774. .13024
                     /gene="grhO3"
     CDS             11774. .13024
                     /gene="grhO3"
                     /codon_start=1
                     /transl_table=11
                     /product="putative cytochrome P450 GrhO3"
                     /protein_id="AAM33670.1"
                     /db_xref="GI:21039505"
                     /translation="MTDTLDEPQTLADGAEDAPAYPVKRTCPYRMPPGYEELREKGPI
                     SRVTLWNGRTAWLVTGNDIGRRLFPDARLSSDVLDPRFPLLAPRIEAQRQQAAAPPLV
                     GVDDPVHARQRRMVLPSFGIRQINALRPEIQKYADDLLDTMLAKGPGVTVDLLTEYAL
                     PMPSAVICMLLGVPYEDHEYFDERSRHVLSSSGEEQAAQAQQAFTEILAYLDDLIVRK
                     QAEPGDTLLDELIARQLEEGKVDRQELAMIATVLLVSGHETTSNMIALSTMALLADPD

  Query Match            39.8%;  Score 40.2;  DB 1;  Length 36905;
  Best Local Similarity  62.4%;  Pred. No. 5.5;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 GTCGAACCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCAC 60
           || ||| |  |||  | ||||||   || || |     || | ||   ||||| ||  |
Db   26435 GTGGAAACCCTGAGTGTGCTGGTGCTGTTGGGCACCGGCCTGGTGGCGGGGGTGCTGTTC 26494

Qy      61 GCCGCTGGCGTCAGCGTCATGCGTCCACTGGCGGAGACGGA 101
           ||||  |  |||||||||||||   | |||   | || |||
Db   26495 GCCGTGGCGGTCAGCGTCATGCCCGCCCTGATCGCGATGGA 26535


RESULT 13
AL646079
LOCUS       AL646079              213050 bp    DNA     linear   BCT 07-DEC-2001
DEFINITION  Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
            segment 4/11.
ACCESSION   AL646079 AL646053
VERSION     AL646079.1  GI:17430956
KEYWORDS    .
SOURCE      Ralstonia solanacearum.
  ORGANISM  Ralstonia solanacearum
            Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
            Ralstonia.
REFERENCE   1  (bases 1 to 213050)
  AUTHORS   Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
            Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
            Chandler,M., Choisne,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
            Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schiex,T.,
            Siguier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
            Weissenbach,J. and Boucher,C.A.
  TITLE     Genome sequence of the plant pathogen Ralstonia solanacearum
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 213050)
  AUTHORS   Boucher,C.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
            Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
            Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
            BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
            Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
            118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
            URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
            Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
            F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
            INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT     Christian.Boucher@toulouse.inra.fr
            http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES             Location/Qualifiers
     source          1. .213050
                     /organism="Ralstonia solanacearum"
                     /strain="GMI1000"
                     /db_xref="taxon:305"
                     /plasmid="megaplasmid"
     gene            100. .1632
                     /gene="RSp0487"
                     /note="RS00365"
     CDS             100. .1632
                     /gene="RSp0487"
                     /function="macromolecule metabolism; macromolecule
                     degradation; degradation of proteins, peptides,
                     glycopeptides"
                     /note="Product confidence : hypothetical
                     Gene name confidence : hypothetical
                     predicted by Codon_usage
                     predicted by Homology
                     predicted by FrameD"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
                     /protein_id="CAD17638.1"
                     /db_xref="GI:17430957"
                     /translation="MDAPFSGRGAVIRRAVGRLLLVLVVGLAVLAVCIAFNTWRHDSR
                     QLDVPPAAPLAVDGAAAAARLGEAIRARTISSFTDAQQNADQFRQLHAMLAARYPKAH
                     AAMQREPVGDFSLLYTWKGSDPSLKPILLMAHQDVVPVAPGTEGDWTEPPFAGVVKDG
                     MVWGRGAWDDKGNLIAQMEAAELLAASGFRPRRTIHFAFGADEEVGGERGAARIAALL
                     KSRGEQLAFVIDEGLLITEGVVPGLTKPAALIGVTEKGFLSVALKLSATPGHSSMPPA
                     PGESAIAMMSAALKHLDDQQLPAGIRGVTREMFETLAPEMRGLNRVALSNLWLFEPLV
                     RKQLQASPSTNAVLQTTTALTIVQAGNKDNVLPGRAEATVNFRLLPGDSASSVIAHVE
                     QAVRSAVPKGHFELAALPGVSEAAPVSPTQSASYQLIGRTVREVFPGTVVAPGLMVGA
                     TDSRHMIGISDHVFRFSPVRARPEDLARFHGTNERISEANLVELIRFYHRLIHQAADA
                     GA"
     gene            complement(1751. .3148)
                     /gene="czcS"
                     /note="RSp0488; RS00366"
     CDS             complement(1751. .3148)
                     /gene="czcS"
                     /EC_number="2.7.3.-"
                     /function="cell processes; protection responses"
                     /note="Product confidence : probable
                     Gene name confidence : probable
                     predicted by Codon_usage
                     predicted by Homology
```

```
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
                HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
                TRANSCRIPTION REGULATOR PROTEIN"
                /protein_id="CAD17639.1"
                /db_xref="GI:17430958"
                /translation="MMRWPASLAARLVLAFVLVAGVMLTGAGTYLYYALNAQLRATDD
                AMLLGKIVQLRHCAKDTQSIDDIVRDARRYTEVVSGHEGLILRVIGPDGRTLVEANPE
                GEPVAAPAAVPPDVLASESALHTWWSRTGIASRAIAAQARLGNHRDTVTVLVAQDGDS
                RAALMASYRRHIFLAVIAGVAAVALLSLWLVRSTLKGLARIGRQAASVVPSQLDTRLT
                LEGAPRELSDLLAALNAMLTRLQDGFARLSQFSADLAHDFRTPLANLIGQTEVTLAHP
                RPAGEYRAVLESSLEEYARLSRMIEDMLFLARADHARVALSVAPLDARHEIQVLVDYF
                EAVAADRDITIQVCGNGRVRADLTLLRRAINNLLDNALRHTPGGARIDVAVQVLPHST
                AIAVRNPGPGIPAASLPLLFDRFYRGDPARANSERSAGLGLAIVQTIMDLHGGTAQAS
                SPPGGLTEFRLIFPA"
gene            complement(3145. .3819)
                /gene="czcR"
                /note="RSp0489; RS00367"
CDS             complement(3145. .3819)
                /gene="czcR"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
                CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
                /protein_id="CAD17640.1"
                /db_xref="GI:17430959"
                /translation="MRLLIVEDEPKAGDYLLKGLTESGFVADLARSGPDGLYQAMEHD
                YDLIVLDVMLPGMDGWQVIRELRRRKTTPVLFLTARDELSDRLKGLELGADDYLVKPF
                AFAELVARVHTILRRGPMRESEVLEIADLTIDTIKRRVTRAGRRIDLTAKEYALLHLL
                ARRTGEVLSRSLISSQVWDVNFDSNTNVVDVAIRRLRAKVDDPFEHKLIHTLRGMGYV
                LDTVAR"
gene            4089. .4394
                /gene="RSp0490"
                /note="RS00368"
CDS             4089. .4394
                /gene="RSp0490"
                /function="miscellaneous; hypothetical/global homology"
                /note="Product confidence : putative
                Gene name confidence : hypothetical
                predicted by Codon_usage
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
                /protein_id="CAD17641.1"
                /db_xref="GI:17430960"
                /translation="MSVRRNCLLPLAAAAAFLACGAATAASVSEQALSPLQVHGARDP
                YTDGARGTRDPYTDGGAQAEATLQASVGPRDPYSDGGMSRSSAPRDRDPYTDGAQAA"
gene            4400. .5725
                /gene="czcC"
                /note="RSp0491; RS00369"
CDS             4400. .5725
                /gene="czcC"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
                RESISTANCE PROTEIN"
                /protein_id="CAD17642.1"
                /db_xref="GI:17430961"
                /translation="MVTHPCIRAAVALSCALLVRSGLAAAPAGEAAAASDTAAPAAQS
                APASVPGAAHERMTLQQAIDRALATHPALAAAGYAVRASEGAVDQASRLRNPEAAFLT
                EDVRRDRATMTAQLNIPLELGGKRAARTRAAELARDAAGQDAAVLAAELRADVTRAFF
                ELSIAQERLRLAQASDALAAQAEAAAVRRVQAGKVSPVEQTKAAVARAGAAIELRDAQ
                ADLLVAQRALSAFWGGAVAVPEADGSAETLPAVPPEAHRVLAAATAAEPVALASTPRA
                LRARLEIEHRQALVNVERSKRVPDVTLSVGAKRDSGANANMAVVGIAVPLPLFDRNQG
                SLLEAQRLADKAAEDYRALRLEQAAALSQDVARLEAARQAVGALRQDVLPGAGRAYDA
                ARIGFEAGKFNFLDVLDAQRTLFQARAQYLAALSRAHQAAAGIDRILGR"
gene            5752. .6966
                /gene="czcB"
                /note="RSp0492; RS00370"
CDS             5752. .6966
                /gene="czcB"
                /function="cell processes; protection responses"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
                predicted by Homology
                predicted by FrameD"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="PROBABLE COBALT-ZINC-CADMIUM RESISTANCE (CATION
                EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
                /protein_id="CAD17643.1"
                /db_xref="GI:17430962"
                /translation="MTLPRQQRRAIALIVALGALAGAALWWGMGARQAAHAESEAEQP
                AAHAEHAQAPKAEAAAEAPHAETIAMTAEAIANAAIGVETAGAAQLRTEVVLPGEIRLN
                DDRTAHVVPRVAGVAEQVAVELGQPVTRGQLLAVLSSPALSDQRSELQAAQQRLALAQ
                EAHAREKRLWEQGIAAAQDYQQARTALQEARIAADNARQKLAAIGAAPAGTALNRFEL
                RAPFDGVVVARHLSQGEAVQAEAAVFTIADLRTVWADFAVTAKDLEAVTTHATATVRA
                TATGTAVQGKVSYVGALLGEQTRSAPARVTLDNPKRAWRPGMFVSVSVESGRVPVPVA
                VRADAVQTVDEKPSVFVPVPAGFEVRPVKTGRSDGTWTEILDGLPAGTRYAAANSYVL
                KAEAGKSADHGH"
gene            6977. .10171
                /gene="czcA"
                /note="RSp0493; RS00371"
CDS             6977. .10171
                /gene="czcA"
                /function="cell processes; transport of small molecules"
                /note="Product confidence : probable
                Gene name confidence : probable
                predicted by Codon_usage
```

Query Match 34.5%; Score 34.8; DB 1; Length 213050;
Best Local Similarity 60.6%; Pred. No. 76;
Matches 57; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy      7 CCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
          | |||||| ||||||||| | ||| |||    |  || |   | | |||    ||   |
Db   9923 CTGATGACGGCGCTGGTTGCCTCGCTCGGCTTCGTGCCGATGGCGCTGGCCACCGGGACC 9982

Qy     67 GGCGTCAGCGTCATGCGTCCACTGGCGGAGACGG 100
          |||| |   ||   |||||| ||||||  |  ||
Db   9983 GGCGCCGAGGTGCAGCGTCCGCTGGCGACGGTGG 10016


RESULT 14
AF009336
LOCUS       AF009336               6286 bp    DNA     linear   BCT 06-OCT-1999
DEFINITION  Streptomyces coelicolor lipase (lipA), and LipR activator (lipR)
            genes, complete cds.
ACCESSION   AF009336
VERSION     AF009336.1  GI:4102168
KEYWORDS    .
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
```

```
REFERENCE   1  (bases 1 to 6286)
  AUTHORS   Valdez,F., Gonzalez-Ceron,G., Kieser,H.M. and Servin-Gonzalez,L.
  TITLE     The Streptomyces coelicolor A3(2) lipAR operon encodes an
            extracellular lipase and a new type of transcriptional regulator
  JOURNAL   Microbiology 145 (Pt 9), 2365-2374 (1999)
  MEDLINE   99445177
   PUBMED   10517589
REFERENCE   2  (bases 1 to 6286)
  AUTHORS   Valdez,F. and Servin-Gonzalez,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-1997) Molecular Biology, Biomedical Research
            Institute. National Autonomous University of Mexico, Apartado
            Postal 70228, Ciudad Universitaria, D.F. 04510, Mexico
FEATURES             Location/Qualifiers
     source          1. .6286
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2) M145"
                     /db_xref="taxon:100226"
     CDS             complement(<1. .584)
                     /note="similar to the E. coli MocA protein"
                     /codon_start=1
                     /transl_table=11
                     /product="orf1"
                     /protein_id="AAD09314.1"
                     /db_xref="GI:4102169"
                     /translation="MRIGLLGTGPWADMAHAPALAGHDDLDFAGVWGRRPDAAEELAQ
                     RHGTRAYADLDALLADVEAVAVALPPAVQAGLAVRAARAGCHLLLDKPLASTVAQGRA
                     VVEAVRTADVASVVFFTTRFQPETEAWITEQSSRDDWFTARAQWLGAVFGGEDSPFAD
                     SPWRREKGALWDVGPHALSVLLPVLGDARRVAAAA"
     gene            975. .1907
                     /gene="lipA"
     CDS             975. .1907
                     /gene="lipA"
                     /note="extracellular lipase"
                     /codon_start=1
                     /transl_table=11
                     /product="lipase"
                     /protein_id="AAD09315.1"
                     /db_xref="GI:4102170"
                     /translation="MQQNDHTHAAPGAARPVLRGVRRRLAAVTAAVARVLVLGTLTGP
                     GAQAADNPYERGPAPTESSIEALRGPYSVADTSVSSLAVTGFGGGTIYYPTSTSDGTF
                     GAVVIAPGFTAYQSSIAWLGPRLASQGFVVFTIDTNTTLDQPDSRGRQLLAALDYLTG
                     RSSVRGRIDSGRLGVMGHSMGGGGTLEAAKSRPSLQAAIPLTPWNLDKSWPEVSTPTL
                     VVGADGDTIAPVASHAEPIYSGLPSSTDRAYLELNNATHFSPNTSNTTIAKYSISWLK
                     RFIDDDTRYEQFLCPLPRESLTIEEYRGNCPHGS"
     gene            2078. .4903
                     /gene="lipR"
     CDS             2078. .4903
                     /gene="lipR"
                     /note="putative transcriptional activator of lipA gene"
                     /codon_start=1
                     /transl_table=11
                     /product="LipR activator"
                     /protein_id="AAD09316.1"
                     /db_xref="GI:4102171"
                     /translation="MTGQTPVRTAVRVHGRSAQRRALRAMLDALPAHGGRLLLAGEPG
                     LGRTTLLQWAARSFRAGPVGHLGPGPDANAHLTAGQLLDTLRAAAVAAPVLVCVDDAH
                     RWDAPARTELARAARLLSTADRVGLLLSVASDRAVGPEFARLPVVHLDPLAPARAAAL
                     LDDLTDGVAAPAVREQLLAEAEGNPALLVALVRRLSPAELRGHRPLPSPLAEAGVLAE
                     AAGGVPTGPSAHARDLLLTTVAAAGEASDGSDADAGLVRDALRRLRPAPGHGPGLGPDP
                     GLGPDPGAEPLPGQLVLAGDRLRFRSPLVRRTVHATAAADRRRAVHRALADALDGGGH
                     RLTALLHRSWSLDGPAPAPASADRLAATAADGAVPASHRLRATAYARAAELTADGARR
                     AERYTAAAEQALLAGRPERARPLLAAARDRAAPDAVRGRAELLRGMTELRDGPVGDAH
                     QSLLLAASLLAADDPAEAALAALAAADAAWAAGDLPACLATLAPEPPLRAEHPSPGSD
                     GPAAARTAAVVHDHRVGMRALLEGRLDRAGAPLGQVVRRSRTGDRPETLLRSAAAALL
                     LGDVDAARSGGARALAAARQLGSDALVPQALEYLAYAELRAGRHPQARTHAEEGLRTA
                     LRAGQRNTAAHHRAVLALASIEEEPDAVARHVAAALDTARRHGLAQAAALAQWAAAR
                     ADLGRGRPFDAADRLGLLVLPGPGRGHFAVWRLAVPCFVEAAVLAGRDEDARRVLTDF
                     ADWAAFGADPQAAAQLARCHALLAPAEHADALYRHALDRHDEAGGDFERARTALLYGK
                     WLRRRRRPREARGRLGTALAGFERCGAGVWAAQARDELRALGAAPRSPGAGAGTPARL
                     GRLTPQQLRIARQVAEGATNREVALTLGVSTRTVDYHLRKVFAALGVRSRLELARMVG
                     QAERAQEAQAAQDARRAGHQEKTGAQP"
     CDS             5024. .5956
                     /note="similar to protein encoded by gene downstream of
                     the Streptomyces exfoliatus M11 lipR gene"
                     /codon_start=1
                     /transl_table=11
                     /product="orf4"
                     /protein_id="AAD09317.1"
                     /db_xref="GI:4102172"
                     /translation="MQHAVRSRAAIRTGLTPVPRPRTPGVTSLIDADALRVLHRAARA
                     LLDDLPDLTDRLMALLREQEPAYRAALTKDPTATWQEATAPCGTACSLLDPRGARDAA
                     RRCSWRIGAARAEQGLPLDALLHAFRLGGSLVWQRLVEETSRAAPEDVRLLVHVAADV
                     WNFVDEHCTLVADAYRQTEWQLGRRRENRVRQLAAGLLDGTGRIADLPEAARALGLPE
                     HGRYVVVAVAGGPPAGPTRPRRRTAGSGSTGTPGRRPTTASSRSGDHDGLPEPHRPEA
                     RRRTPPSPASGSVSAARSTDSPLSGTPVGSRTRP"
BASE COUNT      745 a   2625 c   2239 g    677 t
ORIGIN

  Query Match             34.1%;  Score 34.4;  DB 1;  Length 6286;
  Best Local Similarity   60.9%;  Pred. No. 1.5e+02;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       7 CCGATGACAGCGCTGGTTTCGTCGGTCGAGCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
           |||  ||||  ||||   ||| |||  | | | | |||||  ||||  ||| |||| ||
Db    3737 CCGGAGACACTGCTGCGGTCGGCGGCGGCGGCGCTGCTGCTCGGGGACGTGGACGCAGCC 3796

Qy      67 GGCGTCAGCGTCATGCGTCCACTGGCGGAGAC 98
             ||  | |||     ||  | ||||| | | |
Db    3797 CGCAGCGGCGGGGCCCGGGCGCTGGCCGCGGC 3828


RESULT 15
SCF42/c
LOCUS       SCF42                  34713 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid F42.
ACCESSION   AL137165 AL645882
VERSION     AL137165.2  GI:20520910
KEYWORDS    ABC transporter protein, ATP-binding component; AraC-family
            regulator; branched-chain amino acid ABC transport permease;
            branched-chain amino acid transport ATP-binding protein;
            DNA-binding protein; glycosyl hydrolase; glycosyl transferase;
            hydrolase; integral membrane protein; lipA; lipase; lipoprotein;
            lipR; membrane protein; oxidoreductase; transcriptional activator.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 34713)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
REFERENCE   2  (bases 1 to 34713)
  AUTHORS   Oliver,K. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 34713)
  AUTHORS   Thomson,N.R., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2000) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:6752308.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
```

```
          strand).
          The more significant matches with motifs in the PROSITE database
          are also included but some of these may be fortuitous. The length
          in codons is given for each CDS.
          Usually the highest scoring match found by fasta -o is given for
          CDS which show significant similarity to other CDS in the database.
          The position of possible ribosome binding site sequences are given
          where these have been used to deduce the initiation codon. Gene
          prediction is based on positional base preference in codons using a
          specially developed Hidden Markov Model (Krogh et al., Nucleic
          Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
          of Bibb et al., Gene 30:157-66(1984) as implemented at
          http://www.nih.go.jp/
          jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
          correct initiation codon.  Where possible we choose an initiation
          codon (atg, gtg, ttg or (att)) which is preceded by an upstream
          ribosome binding site sequence (optimally 5-13bp before the
          initiation codon).  If this cannot be identified we choose the most
          upstream initiation codon.
          IMPORTANT: This sequence MAY NOT be the entire insert of the
          sequenced clone.  It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions. Cosmid F42 Lies
          alongside and overlaps with cosmid F15 on the AseI-F genomic
          restriction fragment.
FEATURES             Location/Qualifiers
     source          1. .34713
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid F42"
     misc_feature    2. .892
                     /note="Pfam match to entry PF01315 Ald_Xan_dh_C, Aldehyde
                     oxidase and xanthine dehydrogenase, C terminus, score
                     -294.50, E-value 0.00012"
     gene            complement(1187. .2644)
                     /gene="SCF42.02c."
                     /note="SCO0692"
     CDS             complement(1187. .2644)
                     /gene="SCF42.02c."
                     /note="SCF42.02c, possible membrane protein, len: 485 aa.
                     Rich in the amino acid Ala. Contains possible membrane
                     spanning hydrophobic domains."
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein."
                     /protein_id="CAB69664.1"
                     /db_xref="GI:6752310"
                     /db_xref="SPTREMBL:Q9L2L1"
                     /translation="MTTGSSSSRSPEYAISALDLPTGETTTTTGTARPPGTEAEARAS
                     SARPSPEQESDIWDVDFDDADLPEPEPDPDLPASGLPAPDGPGPRALDADAGPAPAAG
                     ADADAAHGYGDPVGDLVRAAVADRPLDEVVDLITTLERSPQYAQATVDALRAVGVNRS
                     VEDVTRLVGLLTRPPRQSDSADEAIRAAAECRSVEDVTRLMALLHRTPLEPHCGQAAV
                     RAAATGRPVEELVELIGRLAESGHMRPERPDARRLRADRDAADEDDGDGRAPRSAGRG
                     GTSAGADRRGRERRRPPRPFRRGRDRDETRAARDRDRARTARDRDRARAVRDRDKARA
                     VRDRAARRAARGPAWPAWLTVAVLAGCGVAYFPLHRGDASAPAYGVALGLSALCLVLA
                     LLLTVRPAVPMLAAAVVGPAALAAAKLYGSATPSARVTPVTDLTLAPVWVAVAVAVAA
                     SLVALMALCVRVASQDSARRRPARPVAVSSRTAAD"
     RBS             2816. .2820
     gene            2822. .3169
                     /gene="SCF42.03"
                     /note="SCO0693"
     CDS             2822. .3169
                     /gene="SCF42.03"
                     /note="SCF42.03, possible membrane protein, len: 115 aa.
                     Contains a possible N-terminal signal sequence and a
                     possible membrane spanning hydrophobic domain. Rich in Ala
                     and Leu."
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein."
                     /protein_id="CAB69665.1"
                     /db_xref="GI:6752311"
                     /db_xref="SPTREMBL:Q9L2L0"
                     /translation="MHRAPEEGAVDTTTVIVVAVLLAALVLAAAVAVLVRLVRTRRDL
                     RRAGLPTGPRWVFWGAVLYLVLPADLLPDPVYLDDIGVLLLALRSARGATGGRLGPAE
                     RRRTTGPPGDYAA"
     RBS             3280. .3283
     gene            3289. .3480
                     /gene="SCF42.04"
                     /note="SCO0694"
     CDS             3289. .3480
                     /gene="SCF42.04"
                     /note="SCF42.04, unknown, len: 62 aa."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SCF42.04."
                     /protein_id="CAB69666.1"
                     /db_xref="GI:6752312"
                     /db_xref="SPTREMBL:Q9L2K9"
                     /translation="MQPFALNYARPAVELEATTPYVYDSGLQLNVLLDGRVAACDHAL
                     LRELGTTTSTAGSKTHFDD"
     gene            3497. .4483
                     /gene="SCF42.05"
                     /note="SCO0695"
     CDS             3497. .4483
                     /gene="SCF42.05"
                     /note="SCF42.05, unknown, len: 328 aa; similar to
                     TR:Q9KZT0 (EMBL:AL353832) hypothetical protein SCE6.08
                     from Streptomyces coelicolor (326 aa) fasta scores; opt:
                     371, z-score: 407.1, E(): 4.8e-15, 35.629% identity
                     (39.145% ungapped) in 334 aa overlap."
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein SCF42.05."
                     /protein_id="CAB69667.1"
                     /db_xref="GI:6752313"
                     /db_xref="SPTREMBL:Q9L2K8"
                     /translation="MTVLILTSEEDVTADMVVVHLNASGVPVVRLDPADLTDSVALSG
                     EFAHGSFRGHLSSGGRLVSIGGLRSVWVRRPGGAATRAAEPSAWLTEEAGQALYGMLR
                     GSGARWMNQPDAAHRARYKPWQLRLAQRCGLPVPATLITTFPRAAREFAERYPDLVVK
                     PVSGAHPQDPPLAVPTSRVPPEADFSAVAHGPTLLQRRVAKRADIRLTAVGEELLAAR
                     KTALASLDPDEVDVRFAGSGEPWRPAEVPPRVAEGVRAYLRAAGLAYGALDFAEDGDG
                     TWWFLECNQSGQFGFVEVDTGQPIARTIAEWLARPGAADAVEPGGPDAAAVG"
     gene            complement(4823. .5317)
                     /gene="SCF42.06c"
                     /note="SCO0696"
     CDS             complement(4823. .5317)
                     /gene="SCF42.06c"
                     /note="SCF42.06c, possible integral membrane protein, len:
                     164 aa. Contains multiple possible membrane spanning
                     hydrophobic domains."
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein."
                     /protein_id="CAB69668.1"
                     /db_xref="GI:6752314"
                     /db_xref="SPTREMBL:Q9L2K7"
                     /translation="MTKILLSLHVLAAIVAVGPVTVAASMFPAAARRVPVAATVGGAD
                     GAAAGDAGGVGTVRLLHRICRVYAGLGIAVPVLGFATAAAMGVLSDAWLVASITLTAV
                     AAGLLAAFVLPRQEELLEELADGRPVERARTARLAMFTGVFNLLWATVTVLMIVRPGS
                     TTGA"
     gene            5572. .6459
                     /gene="SCF42.07"
                     /note="SCO0697"
                     /pseudo
     CDS             5572. .6459
                     /gene="SCF42.07"
                     /note="SCF42.07, possible AraC-family regulator,
                     pseudogene, len: 309 aa. Similar to several other putative
                     AraC-family transcriptional regulators from Streptomyces
                     coelicolor e.g. TR:CAB61299(EMBL:AL132991) SCF55.29 (311
                     aa) and TR:Q9XA73 (EMBL:AL096822) SCGD3.05 (334 aa).
                     Contains a frameshift after codon 188, following a
                     possible deletion event."
                     /pseudo
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="putative AraC-family regulator (pseudogene)"
     gene            complement(6487. .5852)
                     /gene="SCF42.08c"
                     /note="SCO0698"
     CDS             complement(6487. .5852)
                     /gene="SCF42.08c"
                     /note="SCF42.08c, unknown, len: 121 aa."
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein SCF42.08c."

  Query Match             34.1%;  Score 34.4;  DB 1;  Length 34713;
  Best Local Similarity   60.9%;  Pred. No. 1.2e+02;
  Matches   56;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       7 CCGATGACAGCGCTGGTTTCGTCGGTCGACCCCCCGCTGCGAGGGGTGGTGCACGCCGCT 66
           |||  ||||  ||||   ||| |||  | | | | |||||  ||||  ||| |||| ||
Db   21461 CCGGAGACACTGCTGCGGTCGGCGGCGGCCGCGCTGCTGCTCGGGGACGTGGACGCAGCC 21402

Qy      67 GGCGTCAGCGTCATGCGTCCACTGGCGGACAC 98
             ||  | |||     ||  | ||||| | | |
Db   21401 CGCAGCGGCGGGGCCCGGGCGCTGGCCGCCGC 21370


Search completed: November  5, 2002, 12:21:37
Job time : 620.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850
Perfect score: 101
Sequence: 1 cgtgggtgtcgatcgcggcg.....cggggttcgcggcgcgcgcc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	40.6	40.2	630	17	CNS04HMV	AL291136 Tetraodon
	2	40	39.6	470	12	BF588148	BF588148 FM1_38_C1
	3	39.6	39.2	560	14	BQ621957	BQ621957 fch1c.pk0
c	4	38	37.6	442	10	AW620463	AW620463 sj05e03.y
	5	37.4	37.0	368	9	AU070009	AU070009 AU070009
c	6	37.4	37.0	634	14	BQ472223	BQ472223 HV04M06r
c	7	37.4	37.0	657	14	BQ752048	BQ752048 EST632611
	8	37.2	36.8	309	13	BM500473	BM500473 PAC000000
	9	37	36.6	642	12	BG908086	BG908086 TaLr1165C
	10	37	36.6	696	12	BG908515	BG908515 TaLr1168H
	11	37	36.6	705	12	BG908773	BG908773 TaLr1171B
	12	37	36.6	904	9	AL575298	AL575298 AL575298
c	13	37	36.6	1065	14	BQ669726	BQ669726 AGENCOURT
c	14	36.8	36.4	240	9	AV182926	AV182926 AV182926
c	15	36.8	36.4	300	9	AV181002	AV181002 AV181002
	16	36.8	36.4	491	9	AU198322	AU198322 AU198322
c	17	36.8	36.4	529	9	AU209838	AU209838 AU209838
c	18	36.6	36.2	461	13	BM379667	BM379667 MEST508-G
	19	36.6	36.2	476	10	BE051117	BE051117 za72c12.b
	20	36.6	36.2	488	10	BE051832	BE051832 za88f12.g
c	21	36.6	36.2	494	13	BM380443	BM380443 MEST520-A
c	22	36.6	36.2	505	13	BM379579	BM379579 MEST507-E
c	23	36.6	36.2	507	13	BM380586	BM380586 MEST522-A
	24	36.6	36.2	509	10	BE444499	BE444499 WHE1125_D
c	25	36.6	36.2	550	13	BM382462	BM382462 MEST550-A
c	26	36.6	36.2	586	13	BM380890	BM380890 MEST526-F
c	27	36.6	36.2	693	13	BM379505	BM379505 MEST506-D
c	28	36.6	36.2	1078	14	BQ213179	BQ213179 AGENCOURT
c	29	36.4	36.0	300	9	AV175960	AV175960 AV175960
	30	36.4	36.0	681	14	BU024773	BU024773 QHF6A17.y
	31	36.2	35.8	430	13	BM585410	BM585410 170006872
	32	36.2	35.8	507	13	BM609922	BM609922 170006871
	33	36.2	35.8	550	13	BM614132	BM614132 170006871
	34	36.2	35.8	645	13	BM657053	BM657053 170006874
	35	36.2	35.8	651	13	BM578910	BM578910 170006871
	36	36.2	35.8	652	13	BM614782	BM614782 170006871
	37	36.2	35.8	683	13	BM614560	BM614560 170006871
	38	36.2	35.8	683	13	BM637895	BM637895 170006875
	39	36.2	35.8	693	13	BM592382	BM592382 170006874
	40	36.2	35.8	699	13	BM600538	BM600538 170006870
	41	36.2	35.8	699	13	BM631746	BM631746 170006688
	42	36.2	35.8	704	13	BM578296	BM578296 170006871
	43	36.2	35.8	731	17	AQ745293	AQ745293 HS_5503_A
c	44	36.2	35.8	801	17	CNS03I5B	AL245144 Tetraodon
	45	36.2	35.8	827	12	BG344527	BG344527 HVSMEg000

ALIGNMENTS

RESULT 1
CNS04HMV/c
LOCUS CNS04HMV 630 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 110F12 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL291136
VERSION AL291136.1 GI:8029716
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 630)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 630)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished

```
REFERENCE   3  (bases 1 to 630)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .630
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="110F12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG110DC06SP1-end :
                     PUC-Ori"
BASE COUNT      121 a    198 c    113 g    158 t     40 others
ORIGIN

  Query Match             40.2%;  Score 40.6;  DB 17;  Length 630;
  Best Local Similarity   57.1%;  Pred. No. 8.6;
  Matches   52;  Conservative   10;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           || |  | | |||||||   : | | ||| | ||||: ::|  :::|  :||| | ::||
Db     492 GGCGGAGGTTGCGGCGGCGGMGGTGGCGGTGGAGGTRAWRAARKMRGTGKTGGCGGWRGC 433

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCGGCG 95
           |||| ||| | | | ||||| | | |||| |
Db     432 GGTGGAGGTGGTGGTGGCGGCGGTGGCGGTTG 402


RESULT 2
BF588148
LOCUS       BF588148                 470 bp    mRNA    linear   EST 12-DEC-2000
DEFINITION  FM1_38_C12.b1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF588148
VERSION     BF588148.1  GI:11680472
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 276
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness  and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned  to the greenhouse under natural long days during
                     late April/early May, meristems were  harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT       63 a    151 c    187 g     69 t
ORIGIN

  Query Match             39.6%;  Score 40;  DB 12;  Length 470;
  Best Local Similarity   63.5%;  Pred. No. 12;
  Matches   61;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||| || |  ||| |||    | || || || |||||||| | | |   ||| ||
Db     224 GGTGCCGGTACCGGTGGTGGCGGTGCTGGTGCGGGTGGTGGTGGTGCTGGTGCAGGTGCA 283

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           |||| ||||| | |||||||  |   ||||  || ||
Db     284 GGTGGAGGCGGTGGCGGCGGCGGCGGCGGGACGTGG 319


RESULT 3
BQ621957
LOCUS       BQ621957                 560 bp    mRNA    linear   EST 01-JUL-2002
DEFINITION  fch1c.pk002.18 Conidiobolus cornatus ARSEF 512 Conidiobolus
            coronatus cDNA, mRNA sequence.
ACCESSION   BQ621957
VERSION     BQ621957.1  GI:21649126
KEYWORDS    EST.
SOURCE      Conidiobolus coronatus.
  ORGANISM  Conidiobolus coronatus
            Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
            Ancylistaceae; Conidiobolus.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Freimoser,F.M., Screen,S., Hu,G. and St. Leger,R.J.
  TITLE     EST analysis of genes expressed by the zygomycete pathogen
            Conidiobolus coronatus during optimized secretion of proteins
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Freimoser F. M.
            Department of Entomology
            Universtiy of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Tel: 301 405 16 13
            Fax: 301 314 92 90
            Email: ff34@umail.umd.edu.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Conidiobolus coronatus"
                     /strain="ARSEF 512"
                     /db_xref="taxon:34488"
                     /clone_lib="Conidiobolus cornatus ARSEF 512"
                     /note="Vector: UniZap; Conidiobolus coronatus was grown in
                     minimal medium supplemented with Manduca sexta cuticle and
                     peptone for 18 hours. A cDNA library was constructed in
                     the  unidirectional Lambda vector UniZap."
BASE COUNT      103 a    110 c    181 g    145 t     21 others
ORIGIN

  Query Match             39.2%;  Score 39.6;  DB 14;  Length 560;
  Best Local Similarity   63.8%;  Pred. No. 14;
  Matches   60;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  || |  |||||||| ||||   | | |||    ||||||     |  |  |||||
Db      51 CGGTGGAGGTGATCGCGGTGGTCGCGGAGGCGGTCGTGGTGGTTCCTTCAAGGCTGGAGG 110

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGC 94
              |||||| ||| ||||||||||||   | | |||
Db     111 TCGCGGTGGAGGTGATCGCGGCGGTCGTGGAGGC 144


RESULT 4
AW620463/c
```

```
LOCUS       AW620463               442 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sj05e03.y1 Gm-c1032 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1032-1013 5', mRNA sequence.
ACCESSION   AW620463
VERSION     AW620463.1  GI:7326665
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1099   Std Error: 0.00
            High quality sequence stop: 436.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1032-1013"
                     /clone_lib="Gm-c1032"
                     /tissue_type="Cotyledons of 8-day-old 'Williams'
                     seedlings"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from cotyledons of 8-day-old 'Williams' seedlings which
                     were propagated on paper  towels with distilled water for
                     3 days (etiolated), then greenhouse grown for 5 days in
                     potting soil. The cotyledons were flash-frozen in liquid
                     nitrogen. Stratagene's cDNA Synthesis Kit (catalog number
                     200401) was used to synthesize the cDNA. First-strand
                     synthesis was performed with  5-methyl dCTP, hence the
                     ligated cDNA was hemimethylated. A modification of
                     Stratagene's first-strand synthesis primer was used. An
                     'anchor' nucleotide (V=A, C, or G) was added to the 3' end
                     of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to
                     anchor the primer at the 5' end of the poly(A) tract.
                     After second strand synthesis, the cDNA ends were filled
                     in with cloned Pfu DNA, ligated to EcoRI  adapters and
                     subsequently phosphorylated. The cDNA was then
                     precipitated and redissolved in sterile, RNase-,
                     DNase-free water. The XhoI site  within the first-strand
                     synthesis primer was then restricted by digestion with
                     XhoI from Promega (40U/ul); all XhoI sites in the cDNA
                     would be  protected by their hemimethylated status. The
                     cDNA constructs were size-fractionated with a 500 bp
                     cutoff, using GibcoBRL Life Technologies' cDNA Size
                     Fractionation column. The column eluent was then
                     precipitated, redissolved, and ligated into Stratagene's
                     pBluescript II XR Predigested vector (pBluescript II SK(+)
                     vector that has been digested with EcoRI and XhoI, and
                     phosphorylated by Stratagene). 100% of the white and 87.5%
                     of the blue colonies appear to contain recombinant
                     plasmids with cDNA inserts, based on size (n=28 and 8
                     respectively). This library was  constructed by Dr. Paul
                     Keim and Dr. Virginia Coryell."
BASE COUNT       78 a    180 c     90 g     94 t
ORIGIN

  Query Match             37.6%;  Score 38;  DB 10;  Length 442;
  Best Local Similarity   67.9%;  Pred. No. 32;
  Matches   53;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      20 GGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGC 79
           ||||||  ||| |||| ||  ||||   ||||  | ||||  ||||| | |||||   ||
Db     292 GGTCAAGAGGCGGGAGAAGAGGGTGGAGGCGGCGGAGGAGGGAGCGGGGGAGGCGGCTGC 233

Qy      80 GGCGGGGTTCGCGGCGCG 97
           ||||| |   | || |||
Db     232 GGCGGAGGAGGAGGTGCG 215


RESULT 5
AU070009
LOCUS       AU070009               368 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AU070009 Rice panicle shorter than 3cm Oryza sativa (japonica
            cultivar-group) cDNA clone E31777_9A, mRNA sequence.
ACCESSION   AU070009
VERSION     AU070009.1  GI:5004778
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="E31777_9A"
                     /clone_lib="Rice panicle shorter than 3cm"
                     /dev_stage="shorter than 3cm"
                     /note="Organ: panicle"
BASE COUNT       55 a    116 c    145 g     49 t      3 others
ORIGIN

  Query Match             37.0%;  Score 37.4;  DB 9;  Length 368;
  Best Local Similarity   62.1%;  Pred. No. 44;
  Matches   59;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  | |  || ||| |||| |  | | ||||| || || || |  ||  ||  | || |
Db     146 CGCCGTTCCCGTTCGGGGCGATGCAGGCGCCGGGGCCGGGGGCGTACGTCGGGTTCGACC 205

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCG 95
           | |  | |  |||| | |||||||  |   ||||||
Db     206 ATGGCGCGGCGGCGGTGGCGGCGGCGGCTGCGGCG 240


RESULT 6
BQ472223/c
LOCUS       BQ472223               634 bp    mRNA    linear   EST 30-MAY-2002
DEFINITION  HV04M06r HV Hordeum vulgare cDNA clone HV04M06 5-PRIME, mRNA
            sequence.
```

```
ACCESSION   BQ472223
VERSION     BQ472223.1  GI:21280005
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
            ,A.
  TITLE     Barley ESTs from germinating seeds
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Stein Nils
            Molecular Markers Group, Department Genbank
            Institute of Plant Genetics and Crop Plant Research (IPK)
            Correnssstr. 3, 06466, Gatersleben, Germany
            Tel: 039482-5522
            Fax: 039482-5595
            Email: stein@ipk-gatersleben.de
            Insert Length: 634   Std Error: 0.00
            Plate: 4  row: M  column: 6
            Seq primer: M13rev.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Hordeum vulgare"
                     /cultivar="Barke"
                     /db_xref="taxon:4513"
                     /clone="HV04M06"
                     /clone_lib="HV"
                     /tissue_type="germinating seeds"
                     /dev_stage="germinating seeds (48-96 h)"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
                     cDNA); Site_2: XhoI (3'-end of cDNA); Roots were grown for
                     two days on filter paper at room temperature. Due to a
                     cloning artefact caused by the kit, in most cases the
                     EcoRI site is NOT present, as well as the EcoRIadapter
                     used for cloning. To excise the insert, restriction sites
                     upstream EcoRI should be used (e.g. BamHI, SalI,PstI).
                     NOTE: Also due to the cloning system used Blue/white
                     selection for recombinats is not 100% reliable.Average
                     insert size is 1 kb"
BASE COUNT      127 a    282 c    132 g     93 t
ORIGIN

  Query Match             37.0%;  Score 37.4;  DB 14;  Length 634;
  Best Local Similarity   62.1%;  Pred. No. 43;
  Matches   59;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  ||||  | | | || ||  |  |||  || | ||| || ||| |||| |||| |
Db     176 CGGAGGTGAAGGTGGTGGAGGAGACGGGGTCGGTGGAGGAGGCGATGGCGGCGGTGGTGA 117

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCG 95
           | | || |  || ||| | |  |||| | |||| |
Db     116 AGGTGGCGGTGGAGATGGGGATGGGGATGGCGGAG 82


RESULT 7
BQ752048/c
LOCUS       BQ752048                657 bp    mRNA    linear   EST 18-JUL-2002
DEFINITION  EST632611 DSCT Colletotrichum trifolii cDNA clone pDSCT9-59, mRNA
            sequence.
ACCESSION   BQ752048
VERSION     BQ752048.1  GI:21907453
KEYWORDS    EST.
SOURCE      Colletotrichum trifolii.
  ORGANISM  Colletotrichum trifolii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariomycetes incertae sedis; Phyllachorales; Phyllachoraceae;
            mitosporic Phyllachoraceae; Colletotrichum.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   Samac,D.A., Dickman,M., Town,C.D., Van Aken,S., Utterback,T.,
            Cheung,F. and Fraser,C.M.
  TITLE     ESTs from mycelia of Colletotrichum trifolii race 1
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST632612
            Contact: Deborah A. Samac
            Department of Plant Pathology
            University of Minnesota
            495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
            Tel: 612 625 1243
            Fax: 651 649 5058
            Email: debbys@puccini.crl.umn.edu
            TIGR sequence name: MTSAI59TK More information is available at:
            www.medicago.org
            Seq primer: SKmod (CTA gAA CTA gtg gAT CC).
FEATURES             Location/Qualifiers
     source          1. .657
                     /organism="Colletotrichum trifolii"
                     /strain="race 1"
                     /db_xref="taxon:5466"
                     /clone="pDSCT9-59"
                     /clone_lib="DSCT"
                     /tissue_type="mycelia"
                     /dev_stage="Young, actively growing mycelia (3 days after
                     inoculation) grown in liquid culture (cutin minimal medium
                     containing 2%glucose)."
                     /lab_host="DH5alpha"
                     /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
                     EcoRI; isolate: 2sp2 ; cDNA was prepared from polyA+
                     enriched RNA The cDNA was ligated into Lambda gt11 from
                     Stratagene and packaged using Gigapack packaging extracts.
                     An aliquot of the amplified library was used to transduce
                     E. coli Y1090 and phage DNA was purified from a liquid
                     lysate. The cDNA inserts were gel purified after EcoRI
                     digestion and ligated into pBluescript SK+. Aliquots of
                     the ligation were used to transform E. coli DH5alpha which
                     were plated onto medium with X-gal for selection of
                     recombinants."
BASE COUNT      162 a    215 c    154 g    126 t
ORIGIN

  Query Match             37.0%;  Score 37.4;  DB 14;  Length 657;
  Best Local Similarity   62.1%;  Pred. No. 43;
  Matches   59;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  ||||  | | |||| ||||  |  | |||    ||| ||| ||||||  | ||
Db     593 CGGTGGTGGAGGTGGCGGTGGTCGCTATGGCGGCTGGGGTCGTGGTCGCGGCGGCGGTCG 534

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCG 95
             | ||||   ||| || ||||||  |   ||||||
Db     533 TGGTGGTGGTCGCGGTCACGGCGGTGGCGGCGGCG 499


RESULT 8
BM500473
LOCUS       BM500473                309 bp    mRNA    linear   EST 14-FEB-2002
DEFINITION  PAC000000000531 Pioneer AF-1 array Zea mays cDNA, mRNA sequence.
ACCESSION   BM500473
VERSION     BM500473.1  GI:18659692
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 309)
  AUTHORS   Hunter,B.G., Beatty,M., Singletary,G., Hamaker,B., Larkins,B.A. and
            Jung,R.
  TITLE     Maize opaque endosperm mutations create extensive changes in
            patterns of gene expression
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Jung R
```

```
           Trait and Technology Development, Food and Feed Research
           Pioneer Hi-Bred International, Inc.
           7300 NW 62nd Ave., P.O. Box 1004, Johnston, IA 50131-1004, USA
           Tel: 515 270 5934
           Fax: 515 254 2619
           Email: rudolf.jung@pioneer.com.
FEATURES             Location/Qualifiers
     source          1. .309
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
                     /clone_lib="Pioneer AF-1 array"
                     /note="Vector: pSport1; Site_1: SalI; Site_2: NotI"
BASE COUNT       45 a     89 c    122 g     48 t      5 others
ORIGIN

  Query Match             36.8%;  Score 37.2;  DB 13;  Length 309;
  Best Local Similarity   60.6%;  Pred. No. 49;
  Matches   60;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  |||| | |  |||||||||   | | ||| |  |||  || | |||||||  | |
Db     176 CGGCGGTGGCTACGGCGGCGGTCGCGGCGGCGGCGGCGGTTATGGTGGCGGGCGCCGTGA 235

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCG 99
           |||||   ||||   | |||||     ||||   |||
Db     236 CGGCGGTTATGGCGGCGGTGGCGGCTANGGCGGTCGGCG 274


RESULT 9
BG908086
LOCUS       BG908086                642 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  TaLr1165C02R TaLr1 Triticum aestivum cDNA clone TaLr1165C02 5',
            mRNA sequence.
ACCESSION   BG908086
VERSION     BG908086.1  GI:14315749
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 642)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average inset size is >2.2 kb
            Plate: 165  row: C  column: 02
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .642
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1165C02"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT      138 a    212 c    209 g     82 t      1 others
ORIGIN

  Query Match             36.6%;  Score 37;  DB 12;  Length 642;
  Best Local Similarity   64.7%;  Pred. No. 52;
  Matches   55;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      17 GGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGAT 76
           || ||| || | |  |||||  | || || || |  |  |  ||| | || ||||||||
Db     159 GGAGGTGAAAGAGGAGGAGCCCGCGGCGACCGTGCCCCCGCTGCAGGAGGAGCAGGCGAC 218

Qy      77 CGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||  |   ||||||   |||
Db     219 GGCGGCGGCGGCGGCGGCGAAGGGC 243


RESULT 10
BG908515
LOCUS       BG908515                696 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  TaLr1168H10R TaLr1 Triticum aestivum cDNA clone TaLr1168H10 5',
            mRNA sequence.
ACCESSION   BG908515
VERSION     BG908515.1  GI:14316191
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average inset size is >2.2 kb
            Plate: 168  row: H  column: 10
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1168H10"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT      155 a    224 c    220 g     96 t      1 others
ORIGIN

  Query Match             36.6%;  Score 37;  DB 12;  Length 696;
  Best Local Similarity   64.7%;  Pred. No. 52;
  Matches   55;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      17 GGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGAT 76
           || ||| || | |  |||||  | || || || |  |  |  ||| | || ||||||||
Db     135 GGAGGTGAAAGAGGAGGAGCCCGCGGCGACCGTGCCCCCGCTGCAGGAGGAGCAGGCGAC 194

Qy      77 CGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||  |   ||||||   |||
Db     195 GGCGGCGGCGGCGGCGGCGAAGGGC 219
```

```
RESULT 11
BG908773
LOCUS       BG908773                705 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  TaLr1171B04R TaLr1 Triticum aestivum cDNA clone TaLr1171B04 5',
            mRNA sequence.
ACCESSION   BG908773
VERSION     BG908773.1  GI:14316449
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average inset size is >2.2 kb
            Plate: 171  row: B  column: 04
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1171B04"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT      159 a    226 c    224 g     96 t
ORIGIN

  Query Match             36.6%;  Score 37;  DB 12;  Length 705;
  Best Local Similarity   64.7%;  Pred. No. 52;
  Matches   55;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      17 GGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGAT 76
           || ||| || | |  ||||| | || || || |  |  |  ||| | || ||||||||
Db     149 GGAGGTGAAAGAGGAGGAGCCCGCGGCGACCGTGCCCCCGCTGCAGGAGGAGCAGGCGAC 208

Qy      77 CGCGGCGGGGTTCGCGGCGCGCGGC 101
            ||||||| |   ||||||   |||
Db     209 GGCGGCGGCGGCGGCGGCGAAGGGC 233


RESULT 12
AL575298
LOCUS       AL575298                904 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL575298 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI061YB14 3
            prime, mRNA sequence.
ACCESSION   AL575298
VERSION     AL575298.1  GI:12936329
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 904)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .904
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI061YB14"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT      138 a     87 c    444 g    179 t     56 others
ORIGIN

  Query Match             36.6%;  Score 37;  DB 9;  Length 904;
  Best Local Similarity   53.6%;  Pred. No. 52;
  Matches   52;  Conservative   10;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
             |:|| ||  | | ||:|| :| || |  || || || || | :|: |||:| || |:  |
Db     156 GKGTCTCTGTGGGGGSGGCSATTGAGGGGGCGCKGGSGGGGGKCSGGGGSGGGGGGSGGG 215

Qy      64 CGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCGCGG 100
             |:||  |: |   | || ||||   | | || ||:|
Db     216 GGSTGGGGSGGGSGGGGGGGGGGGGGGTGCCGGGCSG 252


RESULT 13
BQ669726/c
LOCUS       BQ669726                1065 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8208383 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6254498
            5', mRNA sequence.
ACCESSION   BQ669726
VERSION     BQ669726.1  GI:21780560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1065)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2404  row: b  column: 03
            High quality sequence stop: 221.
FEATURES             Location/Qualifiers
     source          1. .1065
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6254498"
                     /clone_lib="NIH_MGC_102"
                     /tissue_type="epidermoid carcinoma, cell line"
```

```
                    /lab_host="DH10B (phage-resistant)"
                    /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                    Site_2: EcoRI; cDNA made by oligo-dT priming.
                    Directionally cloned into EcoRI/XhoI sites using the
                    following 5' adaptor: GGCACGAG(G).  Library constructed
                    by Ling Hong in the laboratory of Gerald M. Rubin
                    (University of  California, Berkeley) using ZAP-cDNA
                    synthesis kit (Stratagene) and Superscript II RT (Life
                    Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      109 a    551 c    254 g    151 t
ORIGIN

  Query Match             36.6%;  Score 37;  DB 14;  Length 1065;
  Best Local Similarity   67.5%;  Pred. No. 51;
  Matches   52;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      24 AATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCG 83
           || |||| |  | ||| || || || ||| |||| | ||| || |  ||||   |||| |
Db     629 AACGGGCGGCGGGAGGGGGGGAGCGAGGGGGTGGGGTAAGGGGGGGGGGCGGGGGCGGGG 570

Qy      84 GGGTTCGCGGCGCGCGG 100
           |||| || ||  || ||
Db     569 GGGTCCGGGGATCGGGG 553


RESULT 14
AV182926/c
LOCUS       AV182926                 240 bp    mRNA    linear   EST 21-JUL-1999
DEFINITION  AV182926 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk644d3 3', mRNA sequence.
ACCESSION   AV182926
VERSION     AV182926.1  GI:5562827
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 240)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .240
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk644d3"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       44 a    115 c     36 g     38 t      7 others
ORIGIN

  Query Match             36.4%;  Score 36.8;  DB 9;  Length 240;
  Best Local Similarity   63.6%;  Pred. No. 61;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||  ||   |||||||| | | |  || |  |||||  || |  | ||||||  | |
Db     223 GGTTCCGGAGGCGGCGGTGGAGGTGGTGGCGGCGGTGGCTATGGAAGCGGTGGAGGATGG 164

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCG 92
           |||| |||| | ||||| | | ||| |
Db     163 GGTGGCGGCGGTGGCGGCAGAGATCGTG 136


RESULT 15
AV181002/c
LOCUS       AV181002                 300 bp    mRNA    linear   EST 21-JUL-1999
DEFINITION  AV181002 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
            embryo Caenorhabditis elegans cDNA clone yk614g12 3', mRNA
            sequence.
ACCESSION   AV181002
VERSION     AV181002.1  GI:5560903
KEYWORDS    EST.
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
            ; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
            Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., Sano
            ,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
            Nomoto,H.
  TITLE     Expressed genes in C.elegans
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Yuji Kohara
            Genome Biology Lab.
            National Institute of Genetics
            Yata 1111, Mishima, Shizuoka 411, Japan
            Tel: 81-559-81-6854
            Fax: 81-559-81-6855
            Email: ykohara@lab.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .300
                     /organism="Caenorhabditis elegans"
                     /strain="N2"
                     /db_xref="taxon:6239"
                     /clone="yk614g12"
                     /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
                     hermaphrodite embryo"
                     /sex="hermaphrodite"
                     /dev_stage="embryo"
BASE COUNT       72 a    120 c     48 g     52 t      8 others
ORIGIN

  Query Match             36.4%;  Score 36.8;  DB 9;  Length 300;
  Best Local Similarity   63.6%;  Pred. No. 60;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||  ||   |||||||| | | |  || |  |||||  || |  | ||||||  | |
Db     278 GGTTCCGGAGGCGGCGGTGGAGGTGGTGGCGGCGGTGGCTATGGAAGCGGTGGAGGATGG 219

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCG 92
           |||| |||| | ||||| | | ||| |
Db     218 GGTGGCGGCGGTGGCGGCAGAGATCGTG 191


Search completed: November  5, 2002, 16:17:58
Job time : 1118.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850
Perfect score: 101
Sequence: 1 cgtgggtgtcgatcgcggcg.....cggggttcgcggcgcgcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	46.4	45.9	15872	10	US-09-861-289-1	Sequence 1, Appli
	2	35	34.7	401	10	US-09-864-761-3936	Sequence 3936, Ap
	3	35	34.7	446	10	US-09-864-761-20699	Sequence 20699, A
c	4	34.6	34.3	2481	10	US-09-894-998-35	Sequence 35, Appl
	5	33.8	33.5	7703	10	US-09-764-887-415	Sequence 415, App
	6	33.8	33.5	7703	10	US-09-764-887-456	Sequence 456, App
	7	33	32.7	1075	10	US-09-864-761-19241	Sequence 19241, A
	8	33	32.7	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	9	32.8	32.5	3715	10	US-09-880-107-2300	Sequence 2300, Ap
c	10	32.6	32.3	681	10	US-09-893-737-327	Sequence 327, App
	11	32.2	31.9	1797	10	US-09-969-708-328	Sequence 328, App
	12	32	31.7	939	10	US-09-815-242-7796	Sequence 7796, Ap
c	13	31.8	31.5	381	10	US-09-735-705-215	Sequence 215, App
c	14	31.8	31.5	381	10	US-09-850-716A-215	Sequence 215, App
c	15	31.8	31.5	381	10	US-09-897-778-215	Sequence 215, App
c	16	31.8	31.5	395	10	US-09-833-790-275	Sequence 275, App
c	17	31.8	31.5	545	10	US-09-833-790-302	Sequence 302, App
c	18	31.8	31.5	954	10	US-09-972-331-26	Sequence 26, Appl
c	19	31.8	31.5	1481	10	US-09-880-107-2100	Sequence 2100, Ap
c	20	31.6	31.3	1018	10	US-09-896-856-7	Sequence 7, Appli
	21	31.4	31.1	264	10	US-09-294-093B-152	Sequence 152, App
	22	31.4	31.1	476	10	US-09-864-761-6	Sequence 6, Appli
	23	31.4	31.1	1035	10	US-09-086-118-28	Sequence 28, Appl
	24	31.4	31.1	1035	10	US-09-835-684-12	Sequence 12, Appl
	25	31.4	31.1	1035	10	US-09-880-371-12	Sequence 12, Appl
	26	31.4	31.1	1035	10	US-09-879-248-16	Sequence 16, Appl
	27	31.4	31.1	1035	10	US-09-770-693-8	Sequence 8, Appli
	28	31.4	31.1	1035	10	US-09-766-348-8	Sequence 8, Appli
	29	31.4	31.1	13842	10	US-09-861-289-30	Sequence 30, Appl
	30	31.4	31.1	36778	10	US-09-861-289-5	Sequence 5, Appli
	31	31.2	30.9	925	10	US-09-823-038A-45	Sequence 45, Appl
	32	31.2	30.9	1729	10	US-09-835-684-10	Sequence 10, Appl
	33	31.2	30.9	1729	10	US-09-880-371-10	Sequence 10, Appl
	34	31.2	30.9	1729	10	US-09-879-248-13	Sequence 13, Appl
c	35	31.2	30.9	3657	10	US-09-728-422-4	Sequence 4, Appli
	36	31.2	30.9	30365	10	US-09-825-414-1	Sequence 1, Appli
	37	31	30.7	372	10	US-09-864-761-22031	Sequence 22031, A
	38	31	30.7	2862	10	US-09-888-615-7	Sequence 7, Appli
c	39	31	30.7	9142	10	US-09-764-877-3562	Sequence 3562, Ap
	40	30.6	30.3	997	10	US-09-452-239-13	Sequence 13, Appl
	41	30.6	30.3	4826	10	US-09-772-304A-1	Sequence 1, Appli
	42	30.6	30.3	5036	10	US-09-918-951-2	Sequence 2, Appli
c	43	30.4	30.1	514	10	US-09-924-035A-175	Sequence 175, App
c	44	30.4	30.1	681	10	US-09-770-149-341	Sequence 341, App
	45	30.4	30.1	759	10	US-09-887-576-810	Sequence 810, App

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 45.9%; Score 46.4; DB 10; Length 15872;
Best Local Similarity 67.7%; Pred. No. 0.0015;
Matches 65; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           || |   |||| ||||| || || ||   |  |||||||||| | |||| |||| |||||
Db   10574 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 10633

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
             |||   | |   |||||| | |||| |  || |||
Db   10634 GTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGC 10669
```

RESULT 2
US-09-864-761-3936
; Sequence 3936, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.

```
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3936
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007249.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
US-09-864-761-3936

  Query Match             34.7%;  Score 55;  DB 10;  Length 401;
  Best Local Similarity   61.5%;  Pred. No. 0.88;
  Matches   56;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           ||||  | | | ||||||    | | ||| |  ||||| | | ||||  |||| |   |
Db     142 GGTGGTGGTGGTGGCGGTGGCGGTGGCGGTGGTGGTGGCGGTGGCGGTAGTGGTGGCGGT 201

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCGGCG 95
```

```
           | ||  |||| | || | || | | | ||||
Db     202 GATGGTGGCGGTGGCAGTGGCGGTGGTGGCG 232


RESULT 3
US-09-864-761-20699
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
;   LENGTH: 446
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC007249.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =  10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.7
;   OTHER INFORMATION: NT HIT: AL161539.2, EVALUE 3.70e+00
```

```
US-09-864-761-20699

  Query Match             34.7%;  Score 35;  DB 10;  Length 446;
  Best Local Similarity   61.5%;  Pred. No. 0.87;
  Matches   56;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           ||||  | | | ||||||    | | ||| |  ||||| | | ||||  |||| |   |
Db      21 GGTGGTGGTGGTGGCGGTGGCGGTGGCGGTGGTGGTGGCGGTGGCGGTAGTGGTGGCGGT 80

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCG 95
           | ||  |||| | || | || | | | ||||
Db      81 GATGGTGGCGGTGGCAGTGGCGGTGGTGGCG 111


RESULT 4
US-09-894-998-35/c
; Sequence 35, Application US/09894998
; Patent No. US20020090610A1
; GENERAL INFORMATION:
;  APPLICANT: Hosken, Nancy Ann
;  APPLICANT:  Craig H. Day
;  APPLICANT:  Davin C. Dillon
;  APPLICANT:  McGowan, Patrick
;  APPLICANT:  Sleath, Paul R.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538
;  CURRENT APPLICATION NUMBER: US/09/894,998
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 2481
;   TYPE: DNA
;   ORGANISM: HSV-2
US-09-894-998-35

  Query Match             34.3%;  Score 34.6;  DB 10;  Length 2481;
  Best Local Similarity   59.8%;  Pred. No. 1;
  Matches   58;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           || |  |   | |||||   | | |  || | ||| || |   ||||  | |||| | ||
Db    1842 GGAGGAGGAGGAGGCGGAGGAGGAGGAGGCGGAGGAGGAGGAGGCGGAGGAGGAGGAGGC 1783

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           || | ||| |   ||||||| |  ||||||  | | |
Db    1782 GGAGGAGGAGGAGGCGGCGGCGACCGCGGCCTGGGAC 1746


RESULT 5
US-09-764-887-415
; Sequence 415, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 415
;   LENGTH: 7703
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-415

  Query Match             33.5%;  Score 33.8;  DB 10;  Length 7703;
  Best Local Similarity   58.4%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  |  | || | | | | |     | | |||||   | || |  |||||||| ||||
Db    7499 CGGAGCGGGCGCTAGAGCCCGAGGCGGAGGCGGAGGCTGAGGCGGGCGCGGGCGGGGAGG 7558

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
             ||| |   ||| |  ||| |||||      |||||| |||
Db    7559 CAGCAGCCGAGGAGGGCGCAGCGGGCCGAAAGGCGCGGGGC 7599


RESULT 6
US-09-764-887-456
; Sequence 456, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 456
;   LENGTH: 7703
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-456

  Query Match             33.5%;  Score 33.8;  DB 10;  Length 7703;
  Best Local Similarity   58.4%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||  |  | || | | | | |     | | |||||   | || |  |||||||| ||||
Db    7499 CGGAGCGGGCGCTAGAGCCCGAGGCGGAGGCGGAGGCTGAGGCGGGCGCGGGCGGGGAGG 7558

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
             ||| |   ||| |  ||| |||||      |||||| |||
Db    7559 CAGCAGCCGAGGAGGGCGCAGCGGGCCGAAAGGCGCGGGGC 7599


RESULT 7
US-09-864-761-19241
; Sequence 19241, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
```

```
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19241
;   LENGTH: 1075
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
;   OTHER INFORMATION: EST_HUMAN HIT: AA739739.1, EVALUE 1.00e+00
;   OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-19
US-09-864-761-19241

  Query Match             32.7%;  Score 33;  DB 10;  Length 1075;
  Best Local Similarity   60.7%;  Pred. No. 2.5;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||   | | | || ||| |    |  || |  |||||||||| | ||  |||| |  ||
Db     103 GGTAGTGGTAGTGGTGGTGATGCTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATAGT 162

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCGG 93
           ||||  || ||| | || || | | | ||
Db     163 GGTGGTGGTGATGGTGGTGGTGGTGGTGG 191


RESULT 8
US-09-864-761-2513
; Sequence 2513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2513
;   LENGTH: 1403
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL078472.1
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

  Query Match             32.7%;  Score 33;  DB 10;  Length 1403;
  Best Local Similarity   60.7%;  Pred. No. 2.5;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||   | | | || ||| |    |  || |  |||||||||| | ||  |||| |  ||
Db     453 GGTAGTGGTAGTGGTGGTGATGCTGGTGGTGGTGGTGGTGATGGTGGTGGTGGTGATAGT 512

Qy      65 GGTGCAGGCGATCGCGGCGGGGGTTCGCGG 93
           ||||  || ||| | || || | | | ||
Db     513 GGTGGTGGTGATGGTGGTGGTGGTGGTGG 541


RESULT 9
US-09-880-107-2300
; Sequence 2300, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
;  APPLICANT: Horne, Darci T.
;  APPLICANT:  Vockley, Joseph G.
;  APPLICANT:  Scherf, Uwe
;  APPLICANT:  Gene Logic, Inc.
```

```
;  TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
;  FILE REFERENCE: 44921-5028-WO
;  CURRENT APPLICATION NUMBER: US/09/880,107
;  CURRENT FILING DATE:  2001-06-14
;  PRIOR APPLICATION NUMBER: US 60/211,379
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/237,054
;  PRIOR FILING DATE: 2000-10-02
;  NUMBER OF SEQ ID NOS: 3950
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2300
;   LENGTH: 3715
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 M23263
US-09-880-107-2300

  Query Match             32.5%;  Score 32.8;  DB 10;  Length 3715;
  Best Local Similarity   64.5%;  Pred. No. 2.7;
  Matches   49;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      20 GGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGC 79
           || || | |   ||| |  |||||| | | ||| |||| |   |||| |  ||||   ||
Db    1843 GGCCAGTTGTATGGACCGTGTGGTGGTGGTGGGGGTGGTGGTGGCGGCGGCGGCGGCGGC 1902

Qy      80 GGCGGGGTTCGCGGCG 95
           ||||| |   ||||||
Db    1903 GGCGGCGGCGGCGGCG 1918


RESULT 10
US-09-893-737-327/c
; Sequence 327, Application US/09893737
; Patent No. US20020110855A1
; GENERAL INFORMATION:
;  APPLICANT: Sheppard, Paul O.
;  APPLICANT:  Presnell, Scott R.
;  TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS
;  FILE REFERENCE: 00-41
;  CURRENT APPLICATION NUMBER: US/09/893,737
;  CURRENT FILING DATE:  2001-06-28
;  PRIOR APPLICATION NUMBER: US 60/215,446
;  PRIOR FILING DATE: 2000-06-30
;  NUMBER OF SEQ ID NOS: 329
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 327
;   LENGTH: 681
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(681)
US-09-893-737-327

  Query Match             32.3%;  Score 32.6;  DB 10;  Length 681;
  Best Local Similarity   63.3%;  Pred. No. 3.2;
  Matches   50;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      16 CGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGA 75
           | |  || |  | || | |||||  || |   |   ||||||||||| |  | ||||| |
Db    461 CTGAAGTAAGCGAGCAGCAGCAGCGGGCGCGGGGCAGCGTGGAGCAGGGCGAGCAGGGGC 402

Qy      76 TCGCGGCGGGGGTTCGCGGC 94
           ||  |||||| ||| |||
Db    401 GCGGCGCGGGGGTCGGGGC 383


RESULT 11
US-09-969-708-328
; Sequence 328, Application US/09969708
; Patent No. US20020102532A1
; GENERAL INFORMATION:
;  APPLICANT: Augustus, Meena
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-70
;  CURRENT APPLICATION NUMBER: US/09/969,708
;  CURRENT FILING DATE:  2001-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,606
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,608
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,425
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 328
;   LENGTH: 1797
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-969-708-328

  Query Match             31.9%;  Score 32.2;  DB 10;  Length 1797;
  Best Local Similarity   59.1%;  Pred. No. 3.8;
  Matches   55;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           || || ||  | |   |  |     || | ||| || |   |||||| || ||||| |
Db    1089 GGCTGGCGGCCACCAAGTGCCGGAAGCGGAAGCTGGAGCGCATCGCGCGCCTGGAGGACA 1148

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGC 96
           |||| || || ||  ||| | |  ||||| ||
Db    1149 AGGTGAAGACGCTCAAGGCCGAGAACGCGGGGC 1181


RESULT 12
US-09-815-242-7796
; Sequence 7796, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7796
;   LENGTH: 939
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
```

```
;   LOCATION: (1)...(939)
US-09-815-242-7796

  Query Match           31.7%;  Score 32;  DB 10;  Length 939;
  Best Local Similarity   62.5%;  Pred. No. 4.4;
  Matches   50;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      22 TCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGG 81
           |||| | ||||| |||||| |||||||| || |  | |   || ||  ||      |   |
Db     485 TCAACGCGCCGGGGCAGGTAGTGATCGCCGGTGCCGCGGCTGCCGTTGAGCGTGCCATCG 544

Qy      82 CGGGGTTCGCGGCGCGCGGC 101
           ||  | |  ||| ||||||
Db     545 AGGCATGCAAGGCACGCGGC 564


RESULT 13
US-09-735-705-215/c
; Sequence 215, Application US/09735705
; Patent No. US20020052329A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Hosken, Nancy
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Li, Samuel X.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.455C14
;  CURRENT APPLICATION NUMBER: US/09/735,705
;  CURRENT FILING DATE:  2000-12-12
;  NUMBER OF SEQ ID NOS:  419
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 215
;   LENGTH: 381
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (17)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (20)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (60)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (61)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (365)
;   OTHER INFORMATION: n=A,T,C or G
US-09-735-705-215

  Query Match           31.5%;  Score 31.8;  DB 10;  Length 381;
  Best Local Similarity   61.4%;  Pred. No. 5.1;
  Matches   51;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      19 CGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCG 78
           || ||  || ||||  ||  || |||   |     |||  |   ||| ||  ||||   |
Db     231 CGCTCCCTGCGCCGCTGCCCGTAGTGCCGGGCTTGGTGTCGGCGGCGCTGAGGGCGGGGG 172

Qy      79 CGGCGGGGTTCGCGGCGCGCGGC 101
           ||||||||   |||||| |||||
Db     171 CGGCGGGGGGGGCGGCGGGCGGC 149


RESULT 14
US-09-850-716A-215/c
; Sequence 215, Application US/09850716A
; Patent No. US20020115139A1
; GENERAL INFORMATION:
;  APPLICANT: Kalos, Michael D.
;  APPLICANT:  McNeill, Patricia D.
;  APPLICANT:  Retter, Marc W.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C15
;  CURRENT APPLICATION NUMBER: US/09/850,716A
;  CURRENT FILING DATE:  2001-05-07
;  NUMBER OF SEQ ID NOS: 440
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 215
;   LENGTH: 381
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (17)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (20)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (60)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (61)
;   OTHER INFORMATION: n=A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: (365)
;   OTHER INFORMATION: n=A,T,C or G
US-09-850-716A-215

  Query Match           31.5%;  Score 31.8;  DB 10;  Length 381;
  Best Local Similarity   61.4%;  Pred. No. 5.1;
  Matches   51;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      19 CGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCG 78
           || ||  || ||||  ||  || |||   |     |||  |   ||| ||  ||||   |
Db     231 CGCTCCCTGCGCCGCTGCCCGTAGTGCCGGGCTTGGTGTCGGCGGCGCTGAGGGCGGGGG 172

Qy      79 CGGCGGGGTTCGCGGCGCGCGGC 101
           ||||||||   |||||| |||||
Db     171 CGGCGGGGGGGGCGGCGGGCGGC 149


RESULT 15
US-09-897-778-215/c
; Sequence 215, Application US/09897778
; Patent No. US20020147143A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Tongtong
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Peckham, David W.
;  APPLICANT:  Fanger, Neil
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.455C16
;  CURRENT APPLICATION NUMBER: US/09/897,778
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 467
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

; SEQ ID NO 215
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 17, 20, 60, 61, 365
; OTHER INFORMATION: n = A,T,C or G
US-09-897-778-215

Query Match 31.5%; Score 31.8; DB 10; Length 381;
Best Local Similarity 61.4%; Pred. No. 5.1;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy      19 CGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCG 78
           || ||  || ||||  ||  || |||   |     |||  |   ||| ||  ||||   |
Db     231 CGCTCCCTGCGCCGCTGCCCGTAGTGCCGGGCTTGGTGTCGGCGGCGCTGAGGGCGCGGG 172

Qy      79 CGGCGGGGTTCGCGGCGCGCGGC 101
           ||||||||   |||||| |||||
Db     171 CGGCGGGGGGGGCGGCGGGCGGC 149
```

Search completed: November 5, 2002, 23:00:14
Job time : 89.2537 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850
Perfect score: 101
Sequence: 1 cgtgggtgtcgatcgcggcg.....cggggttcgcggcgcgcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	2	99.4	98.4	68750	3	US-09-335-409-1	Sequence 1, Appli
	3	99.4	98.4	68750	4	US-09-568-102-1	Sequence 1, Appli
	4	99.4	98.4	68750	4	US-09-567-969-1	Sequence 1, Appli
	5	99.4	98.4	68750	4	US-09-568-480-1	Sequence 1, Appli
	6	99.4	98.4	68750	4	US-09-568-486-1	Sequence 1, Appli
	7	99.4	98.4	68750	4	US-09-568-472-1	Sequence 1, Appli
	8	99.4	98.4	68750	4	US-09-567-899-1	Sequence 1, Appli
	9	46.4	45.9	15872	4	US-09-105-537-1	Sequence 1, Appli
	10	46.4	45.9	43280	2	US-08-804-227C-1	Sequence 1, Appli
	11	43.2	42.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	12	43.2	42.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	13	41.4	41.0	13987	2	US-08-804-227C-13	Sequence 13, Appl
	14	41.4	41.0	44377	2	US-08-804-227C-7	Sequence 7, Appli
	15	41.4	41.0	44377	2	US-08-804-198-1	Sequence 1, Appli
	16	38.8	38.4	20235	1	US-07-642-734C-3	Sequence 3, Appli
	17	38.8	38.4	20235	3	US-08-439-009A-3	Sequence 3, Appli
	18	38.6	38.2	1035	3	US-08-858-003-30	Sequence 30, Appl
	19	38.6	38.2	1035	3	US-09-078-166-30	Sequence 30, Appl
	20	38.6	38.2	1035	4	US-08-997-467-30	Sequence 30, Appl
c	21	37.8	37.4	3196	4	US-09-704-449-1	Sequence 1, Appli
	22	37.2	36.8	33529	4	US-09-144-085-3	Sequence 3, Appli
	23	36.2	35.8	80161	3	US-09-036-987A-1	Sequence 1, Appli
	24	36.2	35.8	80161	4	US-09-370-700-1	Sequence 1, Appli
	25	35.8	35.4	1028	4	US-08-118-200-1	Sequence 1, Appli
	26	35.8	35.4	1028	4	US-08-458-745-1	Sequence 1, Appli
	27	35.6	35.2	1505	1	US-07-915-246-1	Sequence 1, Appli
	28	35.6	35.2	11219	1	US-07-642-734C-1	Sequence 1, Appli
	29	35.6	35.2	11219	3	US-08-439-009A-1	Sequence 1, Appli
	30	35	34.7	888	4	US-08-765-907A-2	Sequence 2, Appli
c	31	35	34.7	2888	4	US-08-765-907A-1	Sequence 1, Appli
	32	34.2	33.9	10642	4	US-09-934-551-3	Sequence 3, Appli
c	33	33.8	33.5	2721	6	5215881-2	Patent No. 5215881
c	34	33.8	33.5	8438	1	US-07-945-283-1	Sequence 1, Appli
c	35	33.6	33.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	36	33.2	32.9	5011	1	US-08-141-893-1	Sequence 1, Appli
c	37	33.2	32.9	5011	1	US-08-463-092B-1	Sequence 1, Appli
c	38	33.2	32.9	5011	1	US-08-463-092B-3	Sequence 3, Appli
c	39	33.2	32.9	5011	2	US-08-462-109A-1	Sequence 1, Appli
c	40	33.2	32.9	5011	2	US-08-462-109A-3	Sequence 3, Appli
c	41	33.2	32.9	5011	2	US-08-460-907B-1	Sequence 1, Appli
c	42	33.2	32.9	5011	2	US-08-460-907B-3	Sequence 3, Appli
c	43	33.2	32.9	5011	3	US-08-463-179A-1	Sequence 1, Appli
c	44	33.2	32.9	5011	3	US-08-463-179A-3	Sequence 3, Appli
c	45	33.2	32.9	5011	3	US-08-461-384B-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18271 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 18330

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   18331 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 18371


RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             98.4%;  Score 99.4;  DB 3;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||| ||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 4
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 6.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             45.9%;  Score 46.4;  DB 4;  Length 15872;
  Best Local Similarity   67.7%;  Pred. No. 0.0058;
  Matches   65;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
          || |   |||| ||||| || || ||   |  |||||||||| | |||| |||| |||||
Db  10574 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 10633

Qy     66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||   | |   ||||||  | |||| |  || |||
```

```
Db  10634 GTGGCCGAGGCGGCGGCGCAGCTCGCCCGACGAGGC 10669


RESULT 10
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             45.9%;  Score 46.4;  DB 2;  Length 43280;
  Best Local Similarity   67.7%;  Pred. No. 0.0057;
  Matches   65;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           || |   |||| ||||| || || ||   |  |||||||||| | |||| |||| |||||
Db  38436 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 38495

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   | |   ||||||  | |||| |  || |||
Db  38496 GTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGC 38531
```

```
RESULT 11
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             42.8%;  Score 43.2;  DB 4;  Length 4403765;
  Best Local Similarity   65.6%;  Pred. No. 0.028;
  Matches   63;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           |||   |||||||||||||||||| ||||    |||||||||| | || |   |  | |||
Db 489502 GTGAGCATCGCGGCGGTCAATGGTCCGGCTTCGGTGGTGATCTCTGGTGCCCACGATGCG 489561

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   |||||||| |   || |    |  |  |||
Db 489562 GTGAGCGCGATCGCTGATCGGCTGCGCGGCCAGGGC 489597


RESULT 12
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             42.8%;  Score 43.2;  DB 4;  Length 4411529;
  Best Local Similarity   65.6%;  Pred. No. 0.028;
  Matches   63;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           |||   |||||||||||||||||| ||||    |||||||||| | || |   |  | |||
Db 488060 GTGAGCATCGCGGCGGTCAATGGTCCGGCTTCGGTGGTGATCTCTGGTGCCCACGATGCG 488119

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   |||||||| |   || |    |  |  |||
Db 488120 GTGAGCGCGATCGCTGATCGGCTGCGCGGCCAGGGC 488155
```

```
RESULT 13
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..13987
US-08-804-227C-13

  Query Match             41.0%;  Score 41.4;  DB 2;  Length 13987;
  Best Local Similarity   63.6%;  Pred. No. 0.079;
  Matches   63;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 GTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCA 61
          |||||||||| | |||||||| |||||||||  |  ||||||| |  |||| |  | |
Db   5635 GTGGGTGTCGGTTGCGGCGGTGAATGGGCCGCGGTCGGTGGTGTTGTCGGGTGATGTGGG 5694

Qy     62 AGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
            | | || || || | |  ||   | |   || |   ||
Db   5695 TGTGCTGGAGTCGGTGGTTGCCTCGCTGATGGGGGATGG 5733


RESULT 14
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             41.0%;  Score 41.4;  DB 2;  Length 44377;
  Best Local Similarity   63.6%;  Pred. No. 0.077;
  Matches   63;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 GTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCA 61
          |||||||||| | |||||||| |||||||||  |  ||||||| |  |||| |  | |
Db   2437 GTGGGTGTCGGTTGCGGCGGTGAATGGGCCGCGGTCGGTGGTGTTGTCGGGTGATGTGGG 2496

Qy     62 AGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
            | | || || || | |  ||   | |   || |   ||
Db   2497 TGTGCTGGAGTCGGTGGTTGCCTCGCTGATGGGGGATGG 2535


RESULT 15
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
```

```
;       STATE:  IN
;       COUNTRY:  USA
;       ZIP:  46285
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  Macintosh 7.0
;       SOFTWARE:  Microsoft Word 5.1
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/804,198
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  CANTRELL, PAUL R.
;       REGISTRATION NUMBER:  36,470
;       REFERENCE/DOCKET NUMBER:  P9113
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  317-276-3885
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  44377 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  350..14002
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  14046..20036
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  20110..31284
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31329..36071
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36155..41830
US-08-804-198-1

  Query Match             41.0%;  Score 41.4;  DB 2;  Length 44377;
  Best Local Similarity   63.6%;  Pred. No. 0.077;
  Matches   63;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 GTGGGTGTCGATCGCGGCGGTCAATGGGCC3GAGCAGGTGGTGATCGCGGGCGTGGAGCA 61
           |||||||||| | |||||||| |||||||||  |  ||||||| |  |||| |  | |
Db    2437 GTGGGTGTCGGTTGCGGCGGTGAATGGGCC3CGGTCGGTGGTGTTGTCGGGTGATGTGGG 2496

Qy      62 AGCGGTGCAGGCGATCGCGGCGGGGTTCGC3GCGCGCGG 100
             | | || || || | |  ||   | |   || |   ||
Db    2497 TGTGCTGGAGTCGGTGGTTGCCTCGCTGAT3GGGGATGG 2535


Search completed: November  5, 2002, 15:06:40
Job time : 1926.74 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850
Perfect score: 101
Sequence: 1 cgtgggtgtcgatcgcggcg.....cggggttcgcggcgcgcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	71989	21	AAA29349	Sorangium cellulos
	2	99.4	98.4	68750	21	AAZ55887	Sorangium cellulos
	3	60.2	59.6	20394	22	AAF24892	Pimaricin biosynth
	4	51.2	50.7	27541	22	AAD17185	Streptomyces nours
	5	51.2	50.7	125401	22	AAD17186	Streptomyces nours
	6	50.6	50.1	31422	22	AAH79278	Streptomyces averm
	7	49.4	48.9	4851	22	AAH52062	Mycobacterium tube
	8	49	48.5	31422	21	AAA92302	S. avermitilis ave
	9	46.4	45.9	15872	18	AAT68715	Streptomyces venez
	10	46.4	45.9	15872	21	AAZ87283	S. venezuelae vep
	11	46.4	45.9	43280	18	AAT80413	Tylactone synthase
	12	44.2	43.8	65140	22	AAD17184	Streptomyces nours
c	13	43.8	43.4	77536	21	AAA14651	Nucleotide sequenc
	14	43.2	42.8	4209	22	AAH51966	Mycobacterium tube
	15	43.2	42.8	4209	22	AAF31625	Mycobacterium tube
	16	43.2	42.8	4260	22	AAF31645	Mycobacterium tube
	17	43.2	42.8	5544	22	AAF24891	Pimaricin biosynth
	18	43.2	42.8	4403765	22	AAI99683	Mycobacterium tube
	19	43.2	42.8	4411529	22	AAI99682	Mycobacterium tube
	20	42.6	42.2	53789	19	AAV21187	Amycolatopsis medi
	21	41.6	41.2	4466	21	AAA14663	Nucleotide sequenc
	22	41.6	41.2	4478	21	AAA14661	Nucleotide sequenc
	23	41.6	41.2	4674	21	AAA14666	Nucleotide sequenc
	24	41.6	41.2	4737	21	AAA14668	Nucleotide sequenc
	25	41.4	41.0	13987	18	AAT80415	Hybrid srmG/tylG O
	26	41.4	41.0	44377	18	AAT78508	Platenolide syntha
	27	41.4	41.0	44377	18	AAT80414	Platenolide syntha
	28	40.4	40.0	3978	21	AAC55785	Type I polyketide
	29	40.4	40.0	12249	21	AAC55840	Complete Mitomycin
	30	40.4	40.0	18331	21	AAC55857	Complete nucleotid
	31	38.8	38.4	4725	21	AAA14665	Nucleotide sequenc
	32	38.8	38.4	29879	14	AAQ46806	eryA region of S.
	33	38.6	38.2	771	23	ABL25385	Drosophila melanog
	34	38.6	38.2	1035	20	AAV99256	DNA encoding an ac
	35	38.6	38.2	1035	21	AAA72538	DNA encoding the A
	36	38.6	38.2	2831	23	ABL25384	Drosophila melanog
	37	38.6	38.2	3138	23	ABL10496	Drosophila melanog
c	38	38.2	37.8	7720	21	AAA53800	Genomic DNA encodi
	39	38	37.6	37856	21	AAA11992	S. cellulosum DNA
c	40	37.8	37.4	3196	24	ABA01139	Thielavia terrestr
	41	37.6	37.2	954	24	ABQ90247	M. capsulatus gene
	42	37.2	36.8	33529	23	AAS17367	DNA sequence of S.
	43	36.6	36.2	1530	21	AAC50661	Arabidopsis thalia
c	44	36.6	36.2	2373	23	AAS65199	DNA encoding novel
	45	36.4	36.0	77536	21	AAA14651	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAA29349
ID AAA29349 standard; DNA; 71989 BP.
XX
AC AAA29349;
XX
DT 12-SEP-2000 (first entry)
XX
DE Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /*tag= c
FT /label= epoA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548

```
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                         domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                         domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                         domain, potentially involved in formation of the
FT                         thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                         domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                         module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                         NRPS module'
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                         module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                         module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                         module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                         module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-14;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18271 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 18330

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  18331 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 18371


RESULT 2
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
```

```
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR          AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR          AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR          AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             98.4%;  Score 99.4;  DB 21;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.6e-14;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 3
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             59.6%;  Score 60.2;  DB 22;  Length 20394;
  Best Local Similarity   76.3%;  Pred. No. 2.2e-05;
  Matches   74;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           | || ||| |||||||||||||||| |||  |  ||| |||||| ||||||  ||| | |
Db   11918 GCGTCTCGCTCGCGGCGGTCAATGGCCCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACG 11977

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           |||||| ||| |||||||||   ||||||| |   ||
Db   11978 CGGTGCTGGCCATCGCGGCGCACTTCGCGGGGGAGGG 12014


RESULT 4
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
```

```
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match           50.7%;  Score 51.2;  DB 22;  Length 27541;
  Best Local Similarity  70.8%;  Pred. No. 0.0024;
  Matches   68;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||| ||||||| || ||||| |||||| ||   ||||||||| | || |  ||| | ||
Db   13813 GGTGGCGATCGCCGCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTGACGAGGAGGC 13872

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
             |   |  ||||||| ||   |||||| |  |||||
Db   13873 CGCCGAAACGATCGCCGCCACGTTCGCCGAACGCGG 13908


RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match           50.7%;  Score 51.2;  DB 22;  Length 125401;
```

```
  Best Local Similarity   70.8%;  Pred. No. 0.0023;
  Matches   68;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||| ||||||| || ||||| ||||||  ||   |||||||||  | ||  |  |||  | ||
Db   13813 GGTGGCGATCGCCGCCGTCAACGGGCCGCAGTCCGTGGTGATCTCCGGTGACGAGGAGGC 13872

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
             |   |  |||||||  ||   ||||||  |  |||||
Db   13873 CGCCGAAACGATCGCCGCCACGTTCGCCGAACGCGG 13908


RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match           50.1%;  Score 50.6;  DB 22;  Length 31422;
  Best Local Similarity   70.1%;  Pred. No. 0.0033;
  Matches   68;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           |||||||| ||||||| || || || || |    |||||| || | ||||  ||| | |
Db   10250 GGGTGTCGGTCGCGGCCGTGAACGGACCCGCCTCGGTGGTGTTCTCCGGCGCCGAGGACG 10309

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
             ||||  |   || ||||   ||||||| | ||||||
Db   10310 AGGTGGGGAACATGGCGGACTGGTTCGCCGAGCGCGG 10346


RESULT 7
AAH52062
ID   AAH52062 standard; DNA; 4851 BP.
XX
AC   AAH52062;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Mycobacterium tuberculosis potential drug target gene SEQ ID 116.
XX
KW   Drug target; growth; organism viability; characterisation; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200135317-A1.
XX
PD   17-MAY-2001.
XX
PF   13-NOV-2000; 2000WO-US31152.
XX
PR   12-NOV-1999;   99US-0165086.
PR   12-NOV-1999;   99US-0165124.
PR   01-FEB-2000; 2000US-0179531.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Eisenberg D,  Rotstein SH,  Marcotte EM;
XX
DR   WPI; 2001-329193/34.
DR   P-PSDB; AAG81211.
XX
PT   Identifying nucleotide or polypeptide sequence for use as drug target,
PT   involves providing algorithm that analyzes a functional relationship
PT   between nucleotide or polypeptide sequences, and comparing the
PT   sequences -
XX
PS   Disclosure; Page 131-133; 207pp; English.
XX
CC   This invention relates to a method for identifying a nucleotide or
CC   polypeptide sequence that may be a drug target, or essential for growth
CC   or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC   represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC   tuberculosis proteins which are potential drug targets. The DNA and
CC   protein sequences are used to illustrate the method of the invention. The
CC   method involves providing an unknown nucleotide or polypeptide sequences,
CC   and comparing it to a number of sequences along with at least one
CC   algorithm capable of analysing a functional relationship between
CC   nucleotide and polypeptide sequences. The method is useful for
CC   characterising the function of nucleic acids and polypeptides that may be
CC   useful as a target for a drug or essential for the growth or viability of
CC   an organism.
XX
SQ   Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

  Query Match           48.9%;  Score 49.4;  DB 22;  Length 4851;
  Best Local Similarity   68.7%;  Pred. No. 0.0066;
  Matches   68;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       2 GTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCA 61
           | | |||  |||||| ||| |||| | ||| ||   |||||||||| | || | | ||
Db    684 GGGTGTGGGGATCGCTGCGATCAACGCGCCCGAATCGGTGGTGATCTCCGGTGCGCAGGC 743

Qy      62 AGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
             ||||   | ||||| ||||   ||||||| ||||  ||
```

```
Db    744 CGCGGCAAATGCGATTGCGGATCGGTTCGCCGCGCAGGG 782


RESULT 8
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             48.5%;  Score 49;  DB 21;  Length 31422;
  Best Local Similarity   69.1%;  Pred. No. 0.0077;
  Matches   67;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
          |||||||| ||||||| || || || || |    |||||||  | | ||||  ||| | |
Db  10250 GGGTGTCGGTCGCGGCCGTGAACGGACCCGCCTCGGTGGTGTCCTCCGGCGCCGAGGACG 10309

Qy     64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
            ||||  |   || ||||   ||||||| | ||||||
Db  10310 AGGTGGGGAACATGGCGGACTGGTTCGCCGAGCGCGG 10346


RESULT 9
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag=  a
FT   CDS             14056..14136
FT                   /*tag=  b
FT   CDS             14148..15827
FT                   /*tag=  c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             45.9%;  Score 46.4;  DB 18;  Length 15872;
  Best Local Similarity   67.7%;  Pred. No. 0.031;
  Matches   65;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
          || |   |||| ||||| || || ||   |  ||||||||| | |||| |||| |||||
Db  10574 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 10633

Qy     66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||   | |   ||||||  | |||| |  || |||
Db  10634 GTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGC 10669
```

```
RESULT 10
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag= b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag= c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match             45.9%;  Score 46.4;  DB 21;  Length 15872;
  Best Local Similarity   67.7%;  Pred. No. 0.031;
  Matches   65;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           || |   |||| ||||| || || ||   |  |||||||||| | |||| |||| |||||
Db   10574 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 10633

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   | |   ||||||  | |||| |  || |||
Db   10634 GTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGC 10669


RESULT 11
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
```

```
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             45.9%;  Score 46.4;  DB 18;  Length 43280;
  Best Local Similarity   67.7%;  Pred. No. 0.03;
  Matches   65;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           || |   |||| ||||| || || ||   |  |||||||||| | |||| |||| |||||
Db   38436 GTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATCTCCGGCGCGGAGGAAGCG 38495

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   | |   ||||||  | |||| |  || |||
Db   38496 GTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGC 38531


RESULT 12
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             43.8%;  Score 44.2;  DB 22;  Length 65140;
  Best Local Similarity   68.5%;  Pred. No. 0.092;
  Matches   61;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           ||  ||||||| ||  |||| ||||||  |     ||||  |  |||| |||||| | ||
Db   40799 GTAGCGATCGCCGCCATCAACGGGCCGAACGCACTGGTCGTATCGGGTGTGGAGGATGCC 40858

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGC 94
           |     | |||||  |||| ||||||||||
Db   40859 GCCGTCGAGATCGGGGCGCGGTTCGCGGC 40887
```

```
RESULT 13
AAA14651/c
ID   AAA14651 standard; DNA; 77536 BP.
XX
AC   AAA14651;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW   FK-506 binding protein; polyketide compound; transplant rejection;
KW   graft-versus-host disease; uveitis; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (412..1836)
FT                   /*tag= a
FT                   /note= "fkbW gene"
FT   CDS             complement (2020..3579)
FT                   /*tag= b
FT                   /note= "fkbV gene"
FT   CDS             3969..4496
FT                   /*tag= c
FT                   /note= "fkbR2 gene'
FT   CDS             complement (4595..5488)
FT                   /*tag= d
FT                   /note= "fkbR1 gene"
FT   CDS             5601..6818
FT                   /*tag= e
FT                   /note= "fkbE gene"
FT   CDS             6808..8052
FT                   /*tag= f
FT                   /note= "fkbF gene"
FT   CDS             8156..8824
FT                   /*tag= g
FT                   /note= "fkbG gene"
FT   CDS             complement (9122..9883)
FT                   /*tag= h
FT                   /note= "fkbH gene"
FT   CDS             complement (9894..10994)
FT                   /*tag= i
FT                   /note= "fkbI gene"
FT   CDS             complement (10987..11247)
FT                   /*tag= j
FT                   /note= "fkbJ gene"
FT   CDS             complement (11244..12092)
FT                   /*tag= k
FT                   /note= "fkbK gene"
FT   CDS             complement (12113..13150)
FT                   /*tag= l
FT                   /note= "fkbL gene"
FT   CDS             complement (13212..13988)
FT                   /*tag= m
FT                   /note= "fkbC gene"
FT   misc_feature    complement (13452..13662)
FT                   /*tag= n
FT                   /note= "ACP6"
FT   misc_feature    complement (13761..14394)
FT                   /*tag= o
FT                   /note= "KR6"
FT   misc_feature    complement (14517..15294)
FT                   /*tag= p
FT                   /note= "ER6"
FT   misc_feature    complement (15438..16587)
FT                   /*tag= q
FT                   /note= "dehydratase domain (DH) 6"
FT   misc_feature    complement (16587..17820)
FT                   /*tag= r
FT                   /note= "acyltransferase domain (AT) 6"
FT   misc_feature    complement (17820..19053)
FT                   /*tag= s
FT                   /note= "KS6"
FT   misc_feature    complement (19116..19326)
FT                   /*tag= t
FT                   /note= "ACP5"
FT   misc_feature    complement (19464..20097)
FT                   /*tag= u
FT                   /note= "KR5"
FT   misc_feature    complement (20241..21420)
FT                   /*tag= v
FT                   /note= "DH5"
FT   misc_feature    complement (21420..22653)
FT                   /*tag= w
FT                   /note= "AT5"
FT   misc_feature    complement (22653..23892)
FT                   /*tag= x
FT                   /note= "KS5"
FT   CDS             complement (23992..46573)
FT                   /*tag= y
FT                   /note= "fkbB gene"
FT   misc_feature    complement (24163..24373)
FT                   /*tag= z
FT                   /note= "ACP4"
FT   misc_feature    complement (24997..26146)
FT                   /*tag= aa
FT                   /note= "DH4 (inactive)"
FT   misc_feature    complement (26146..27430)
FT                   /*tag= ab
FT                   /note= "AT4"
FT   misc_feature    complement (27430..28684)
FT                   /*tag= ac
FT                   /note= "KS4"
FT   misc_feature    complement (28750..28960)
FT                   /*tag= ad
FT                   /note= "ACP3"
FT   misc_feature    complement (29092..29740)
FT                   /*tag= ae
FT                   /note= "KR3"
FT   misc_feature    complement (29869..31018)
FT                   /*tag= af
FT                   /note= "DH3 (inactive)"
FT   misc_feature    complement (31018..32185)
FT                   /*tag= ag
FT                   /note= "AT3"
FT   misc_feature    complement (32185..33439)
FT                   /*tag= ah
FT                   /note= "KS3"
FT   misc_feature    complement (33505..33715)
FT                   /*tag= ai
FT                   /note= "ACP2"
FT   misc_feature    complement (33823..34480)
FT                   /*tag= aj
FT                   /note= "KR2"
FT   misc_feature    complement (34606..35749)
FT                   /*tag= ak
FT   misc_feature    complement (28750..28960)
FT                   /*tag= al
FT                   /note= "DH2 (inactive)"
FT   misc_feature    complement (35749..37144)
FT                   /*tag= am
FT                   /note= "AT2"
FT   misc_feature    complement (37145..38296)
FT                   /*tag= an
FT                   /note= "KS2"
FT   misc_feature    complement (38371..38581)
FT                   /*tag= ao
FT                   /note= "ACP1"
FT   misc_feature    complement (38677..39307)
```

```
FT                   /*tag=  ap
FT                   /note= "KR1"
FT   misc_feature    complement (39442..40609)
FT                   /*tag=  aq
FT                   /note= "DH1"
FT   misc_feature    complement (40609..41842)
FT                   /*tag=  ar
FT                   /note= "AT1"
FT   misc_feature    complement (41842..43093)
FT                   /*tag=  as
FT                   /note= "KS of extender module 1 (KS1)"
FT   misc_feature    complement (43144..43660)
FT                   /*tag=  at
FT                   /note= "ACP of loading domain"
FT   misc_feature    complement (43777..44629)
FT                   /*tag=  au
FT                   /note= "ER of loading domain"
FT   misc_feature    complement (44974..46573)
FT                   /*tag=  av
FT                   /note= "CoA ligase of laoding domain"
FT   CDS             46754..47788
FT                   /*tag=  aw
FT                   /note= "fkbO gene"
FT   CDS             47785..52272
FT                   /*tag=  ax
FT                   /note= "fkbP gene"
FT   CDS             52275..71465
FT                   /*tag=  ay
FT                   /note= "fkbA gene"
FT   misc_feature    52362..53576
FT                   /*tag=  az
FT                   /note= "KS7"
FT   misc_feature    53577..54716
FT                   /*tag=  bb
FT                   /note= "AT7"
FT   misc_feature    54717..55871
FT                   /*tag=  ba
FT                   /note= "DH7"
FT   misc_feature    56019..56819
FT                   /*tag=  bc
FT                   /note= "ER7"
FT   misc_feature    56943..57575
FT                   /*tag=  bd
FT                   /note= "KR7"
FT   misc_feature    57710..57920
FT                   /*tag=  be
FT                   /note= "ACP7"
FT   misc_feature    57990..59243
FT                   /*tag=  bf
FT                   /note= "KS8"
FT   misc_feature    59244..60398
FT                   /*tag=  bg
FT                   /note= "AT8"
FT   misc_feature    60399..61412
FT                   /*tag=  bh
FT                   /note= "DH8 (inactive)"

  Query Match             43.4%;  Score 43.8;  DB 21;  Length 77536;
  Best Local Similarity   69.0%;  Pred. No. 0.11;
  Matches   60;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           | |||  |||||| ||||||||  ||||| | |  | | ||| || | || |  ||| | |
Db   31536 GCGTGGAGATCGCCGCGGTCAACGGGCCCGCGTCGATCGTGCTCTCCGGTGACGAGCAGG 31477

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCG 90
           ||||||  |   |||||||  || |||
Db   31476 CGGTGCTCGACGTCGCGGCCCGGCTCG 31450


RESULT 14
AAH51966
ID   AAH51966 standard; DNA; 4209 BP.
XX
AC   AAH51966;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Mycobacterium tuberculosis potential drug target gene SEQ ID 20.
XX
KW   Drug target; growth; organism viability; characterisation; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200135317-A1.
XX
PD   17-MAY-2001.
XX
PF   13-NOV-2000; 2000WO-US31152.
XX
PR   12-NOV-1999;   99US-0165086.
PR   12-NOV-1999;   99US-0165124.
PR   01-FEB-2000; 2000US-0179531.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Eisenberg D,  Rotstein SH,  Marcotte EM;
XX
DR   WPI; 2001-329193/34.
DR   P-PSDB; AAG81115.
XX
PT   Identifying nucleotide or polypeptide sequence for use as drug target,
PT   involves providing algorithm that analyzes a functional relationship
PT   between nucleotide or polypeptide sequences, and comparing the
PT   sequences -
XX
PS   Disclosure; Page 60-62; 207pp; English.
XX
CC   This invention relates to a method for identifying a nucleotide or
CC   polypeptide sequence that may be a drug target, or essential for growth
CC   or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC   represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC   tuberculosis proteins which are potential drug targets. The DNA and
CC   protein sequences are used to illustrate the method of the invention. The
CC   method involves providing an unknown nucleotide or polypeptide sequences,
CC   and comparing it to a number of sequences along with at least one
CC   algorithm capable of analysing a functional relationship between
CC   nucleotide and polypeptide sequences. The method is useful for
CC   characterising the function of nucleic acids and polypeptides that may be
CC   useful as a target for a drug or essential for the growth or viability of
CC   an organism.
XX
SQ   Sequence 4209 BP; 751 A; 1067 C; 1412 G; 979 T; 0 other;

  Query Match             42.8%;  Score 43.2;  DB 22;  Length 4209;
  Best Local Similarity   65.6%;  Pred. No. 0.17;
  Matches   63;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           |||   |||||||||||||||| ||||    |||||||||| | || |   |  | |||
Db   2332 GTGAGCATCGCGGCGGTCAATGGTCCGGCTTCGGTGGTGATCTCTGGTGCCCACGATGCG 2391

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   |||||||| |   || |    |  |  |||
Db   2392 GTGAGCGCGATCGCTGATCGGCTGCGCGGCCAGGGC 2427


RESULT 15
AAF31625
ID   AAF31625 standard; DNA; 4209 BP.
XX
AC   AAF31625;
XX
DT   09-APR-2001  (first entry)
XX
DE   Mycobacterium tuberculosis pks (Rv0405) gene.
```

```
XX
KW   Mycobacterium tuberculosis; attenuated microorganism;
KW   signature tagged transposon mutant; mutant library;
KW   mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW   vaccine; pks; Rv0405; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200102555-A1.
XX
PD   11-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-IB00950.
XX
PR   06-JUL-1999;   99US-0142982.
PR   08-JUL-1999;   99US-0142833.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Guilhot C,  Camacho L;
XX
DR   WPI; 2001-091804/10.
DR   P-PSDB; AAB66459.
XX
PT   Screening a mutant library for mutants unable to grow under specific
PT   conditions and for identifying loci involved in pathogenicity,
PT   comprises using signature tagged transposon mutagenesis -
XX
PS   Example 8; Page 93-94; 159pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   screening a library of mutants. The method comprises constructing a
CC   library with insertions in genes and/or regulatory regions of the
CC   organisms of interest, where the insertion contains a tag and/or a
CC   transposon associated with a tag. The mutants are identified by
CC   hybridisation of the tags to known sequences. The method is useful for
CC   treating an individual suffering from a mycobacterial infection,
CC   suspected of being infected with a Mycobacterium, or having been
CC   exposed to an infectious Mycobacterium. It is also useful for
CC   identifying and isolating mutants of actinomycetales and for identifying
CC   compounds that have antibiotic activity. The method is used to identify
CC   mutants of microorganisms, preferably an actinomycetales, such as
CC   M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC   M. paratuberculosis, that is unable to grow under specific conditions.
CC   It is especially useful for identifying loci involved in pathogenicity.
CC   It is useful in constructing vaccines. The method can be used to screen
CC   multiple libraries concurrently. It can screen libraries of different
CC   organisms or different strains of the same organism. The present
CC   sequence was isolated from an attenuated mutant of Mycobacterium
CC   tuberculosis. The sequence has been disrupted by the insertion of the
CC   IS 1096 transposon.
XX
SQ   Sequence 4209 BP; 751 A; 1067 C; 1412 G; 979 T; 0 other;

  Query Match             42.8%;  Score 43.2;  DB 22;  Length 4209;
  Best Local Similarity   65.6%;  Pred. No. 0.17;
  Matches   63;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       6 GTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCG 65
           |||   |||||||||||||||| ||||    |||||||||| | || |   |  | |||
Db    2332 GTGAGCATCGCGGCGGTCAATGGTCCGGCTTCGGTGGTGATCTCTGGTGCCCACGATGCG 2391

Qy      66 GTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||   |||||||| |   || |    |  |  |||
Db    2392 GTGAGCGCGATCGCTGATCGGCTGCGCGGCCAGGGC 2427


Search completed: November  5, 2002, 13:06:48
Job time : 249.215 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_22750_22850
Perfect score: 101
Sequence: 1 cgtgggtgtcgatcgcggcg.....cggggttcgcggcgcgcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	71989	6	AR172664	AR172664 Sequence
3		99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4		99.4	98.4	68750	6	AR193029	AR193029 Sequence
5		99.4	98.4	68750	6	AR199551	AR199551 Sequence
6		99.4	98.4	68750	6	AR199559	AR199559 Sequence
7		99.4	98.4	68750	6	AR199567	AR199567 Sequence
8		99.4	98.4	68750	6	AR201097	AR201097 Sequence
9		99.4	98.4	68750	6	AR208671	AR208671 Sequence
10		80.8	80.0	16124	1	AX024384	AX024384 Sequence
11		80.8	80.0	16124	6	AX024277	AX024277 Sequence
12		60.2	59.6	20394	1	SNA132222	AJ132222 Streptomy
13		60.2	59.6	20394	6	AX067996	AX067996 Sequence
14	c	60.2	59.6	84985	1	SNA278573	AJ278573 Streptomy
15		55.4	54.9	3900	1	AF262754	AF262754 Amycolato
16		53.8	53.3	113193	1	AF357202	AF357202 Streptomy
17		51.2	50.7	27541	6	AX211706	AX211706 Sequence
18		51.2	50.7	123580	1	AF263912	AF263912 Streptomy
19		51.2	50.7	125401	6	AX211739	AX211739 Sequence
20		50.6	50.1	31422	6	E38021	E38021 Avermectin
21	c	50.6	50.1	39314	1	SGR300302	AJ300302 Streptomy
22	c	50.6	50.1	64957	1	AB032367	AB032367 Streptomy
23		50	49.5	78210	1	AB070949	AB070949 Streptomy
24	c	49.4	48.9	16096	1	AE007124	AE007124 Mycobacte
25	c	49.4	48.9	24292	1	MTCY24G1	Z83858 Mycobacteri
26		49.4	48.9	31175	1	MTU00024	U00024 Mycobacteri
27		49.4	48.9	31176	1	MSGY2	AD000009 Mycobacte
28	c	49.2	48.7	26195	1	SC1G7	AL591083 Streptomy
29	c	46.8	46.3	49736	1	AF319998	AF319998 Stigmatel
30		46.6	46.1	66808	1	SAU421825	AJ421825 Stigmatel
31		46.4	45.9	17101	1	AB017641	AB017641 Micromono
32		46.4	45.9	43280	1	SFU78289	U78289 Streptomyce
33		44.2	43.8	65140	6	AX211705	AX211705 Sequence
34		44	43.6	344050	1	MLEPRTN1	AL583917 Mycobacte
35	c	43.8	43.4	77534	1	AF235504	AF235504 Streptomy
36	c	43.4	43.0	18023	1	AX024392	AX024392 Sequence
37	c	43.4	43.0	18023	6	AX024285	AX024285 Sequence
38		43.2	42.8	4209	6	AX069054	AX069054 Sequence
39		43.2	42.8	5544	1	SNA132221	AJ132221 Streptomy
40		43.2	42.8	5544	6	AX067994	AX067994 Sequence
41		43.2	42.8	17798	1	AE006946	AE006946 Mycobacte
42		43.2	42.8	22570	1	MTCY4D9	Z84725 Mycobacteri
43		43.2	42.8	84985	1	SNA278573	AJ278573 Streptomy
44	c	43.2	42.8	107379	1	SHGCPIR	X86780 S.hygroscop
45		43	42.6	104326	1	AB070940	AB070940 Streptomy

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453555"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPCVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYERPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
gene            6260. .10492
                /gene="epoB"
CDS             6260. .10492
                /gene="epoB"
                /note="non ribosomal peptide synthetase; epothilone PKS
                module 1"
                /codon_start=1
                /transl_table=11
                /product="EpoB"
                /protein_id="AAF62881.1"
                /db_xref="GI:7453558"
                /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene            10639. .16137
                /gene="epoC"
CDS             10639. .16137
                /gene="epoC"
                /note="type I polyketide synthase; epothilone PKS module
                2"
                /codon_start=1
                /transl_table=11
                /product="EpoC"
                /protein_id="AAF62882.1"
                /db_xref="GI:7453559"
                /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWKGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match           100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-09;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18271 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 18330

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  18331 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 18371


RESULT 2
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others

ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-09;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18271 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 18330

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  18331 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 18371


RESULT 3
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
```

/db_xref="GI:6724239"
/translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
IGIVMTQSAGFLFSGALERFWRSVYVATESA"
CDS 3415. .5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26906.1"
/db_xref="GI:6724240"
/translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
IQAILRASRDHDLLVIGAESPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
CDS complement(5612. .5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
QHTRLDSRQAENALLSTIAYTDRE"
CDS 6226. .6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene 7610. .11875
/gene="epoA"
CDS 7610. .11875
/gene="epoA"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRPLERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature 7640. .8920
/gene="epoA"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
misc_feature 9236. .10201
/gene="epoA"
/note="Region: acetate acyltransferase"
misc_feature 10529. .11428
/gene="epoA"
/note="Region: enoyl reductase"
misc_feature 11549. .11764
/gene="epoA"
/note="Region: acyl carrier protein"
gene 11872. .16104
/gene="epoP"
CDS 11872. .16104
/gene="epoP"
/note="EPOS P; contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene 16251. .21749
/gene="epoB"
CDS 16251. .21749
/gene="epoB"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 98.4%; Score 99.4; DB 1; Length 68750;

```
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 4
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 5
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 6
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 7
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy     61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983
```

```
RESULT 8
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 9
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 4.7e-09;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23883 CGTCGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 23942

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   23943 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 23983


RESULT 10
AX024384
LOCUS       AX024384               16124 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024384
VERSION     AX024384.1  GI:10184588
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             80.0%;  Score 80.8;  DB 1;  Length 16124;
  Best Local Similarity   88.0%;  Pred. No. 1.2e-05;
  Matches   88;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||| ||||||||||||||| |||||||||||||||||||||||||||||||||  |||
Db    1768 CGTTGGTGTCGATCGCGGCAGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGCCGAGA 1827

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           ||   ||||||  |||||||||||| |||||||||||| ||
Db    1828 AATTCGTGCAGCAGATCGCGGCGGCGTTCGCGGCGCGGGG 1867


RESULT 11
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             80.0%;  Score 80.8;  DB 6;  Length 16124;
  Best Local Similarity   88.0%;  Pred. No. 1.2e-05;
  Matches   88;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       1 CGTGGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGC 60
           ||| ||||||||||||||| |||||||||||||||||||||||||||||||||  |||
Db    1768 CGTTGGTGTCGATCGCGGCAGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGCCGAGA 1827

Qy      61 AAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           ||   ||||||  |||||||||||| |||||||||||| ||
Db    1828 AATTCGTGCAGCAGATCGCGGCGGCGTTCGCGGCGCGGGG 1867


RESULT 12
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
```

KEYWORDS pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
AUTHORS Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
TITLE The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes
JOURNAL J. Biol. Chem. 274 (15), 10133-10139 (1999)
MEDLINE 99214571
PUBMED 10187796
REFERENCE 2 (bases 1 to 20394)
AUTHORS Aparicio,J.F.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av. Real, 1-Leon, 24006, SPAIN

FEATURES Location/Qualifiers
source 1. .20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
gene 1. .20394
/gene="pimS1"
CDS 1. .20394
/gene="pimS1"
/function="pimaricin biosynthesis"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="CAB41041.1"
/db_xref="GI:4678703"
/translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPERGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG

```
                    LIRTPGRRTAAAATEGDVAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                    DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                    VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                    DEDEDFESASDDEVFDLLDNELGLS"
BASE COUNT      2611 a    7486 c    7397 g    2900 t
ORIGIN

  Query Match           59.6%;  Score 60.2;  DB 1;  Length 20394;
  Best Local Similarity  76.3%;  Pred. No. 0.051;
  Matches   74;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCCGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
          | || ||| |||||||||||||||| |||  |  ||| |||||| ||||||  ||| | |
Db   11918 GCGTCTCGCTCGCGGCGGTCAATGGCCCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACG 11977

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           |||||| ||| |||||||||   ||||||| |   ||
Db   11978 CGGTGCTGGCCATCGCGGCGCACTTCGCGGGGGAGGG 12014


RESULT 13
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68042"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24642.1"
                     /db_xref="GI:12329807"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHFVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHDALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRHGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
```

```
                     HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                     LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                     DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                     VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                     DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a   7486 c   7397 g   2900 t
ORIGIN

  Query Match             59.6%;  Score 60.2;  DB 6;  Length 20394;
  Best Local Similarity   76.3%;  Pred. No. 0.051;
  Matches   74;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           | || ||| |||||||||||||||| |||  |  ||| |||||| ||||||  ||| | |
Db   11918 GCGTCTCGCTCGCGGCGGTCAATGGCCCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACG 11977

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           |||||| ||| |||||||||   ||||||| |   ||
Db   11978 CGGTGCTGGCCATCGCGGCGCACTTCGCGGGGGAGGG 12014


RESULT 14
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
gene                 2190. .3566
                     /gene="pimK"
CDS                  2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
gene                 complement(3662. .9736)
                     /gene="pimS4"
CDS                  complement(3662. .9736)
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                     REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                     DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                     GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                     AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                     LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                     MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
```

```
                SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDTAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772.  15198)
                /gene="pimS3"
CDS             complement(9772.  15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPESAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGQADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGIYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVIRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVEHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQGRDPDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFPQGTGARRADLPTYAFQRRRYW
                PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match             59.6%;  Score 60.2;  DB 1;  Length 84985;
  Best Local Similarity   76.3%;  Pred. No. 0.037;
  Matches   74;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       4 GGGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAG 63
           | || ||| |||||||||||||||| |||  |  ||| |||||| ||||||  ||| | |
Db   69235 GCGTCTCGCTCGCGGCGGTCAATGGCCCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACG 69176

Qy      64 CGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 100
           |||||| ||| |||||||||   ||||||| |   ||
Db   69175 CGGTGCTGGCCATCGCGGCGCACTTCGCGGGGGAGGG 69139


RESULT 15
AF262754
LOCUS       AF262754               3900 bp    DNA     linear   BCT 27-SEP-2000
DEFINITION  Amycolatopsis mediterranei DSM 46095 rifamycin biosynthetic gene
            cluster, partial sequence.
ACCESSION   AF262754
VERSION     AF262754.1  GI:10312147
KEYWORDS    .
SOURCE      Amycolatopsis mediterranei.
  ORGANISM  Amycolatopsis mediterranei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Amycolatopsis.
REFERENCE   1  (bases 1 to 3900)
  AUTHORS   Lal,R. and Kaur,H.
  TITLE     Cloning and partial characterization of presumptive rifamycin
            biosynthetic gene cluster from the actinomycete Amycolatopsis
            mediterranei DSM 46095
  JOURNAL   Microbiol. Res. (2000) In press
REFERENCE   2  (bases 1 to 3900)
  AUTHORS   Lal,R. and Kaur,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Zoology, University of Delhi, Delhi 110007,
            India
FEATURES             Location/Qualifiers
     source          1. .3900
                     /organism="Amycolatopsis mediterranei"
                     /strain="DSM 46095"
                     /db_xref="taxon:33910"
     CDS             106. .>3900
                     /note="similar to Amycolatopsis mediterranei S699
                     rifamycin biosynthetic gene cluster; contains module 2
                     ACP, all of module 3 including KS, AT, and ACP, as well as
                     part of module 4"
                     /codon_start=1
                     /transl_table=11
                     /product="rifamycin polyketide synthase"
                     /protein_id="AAG16689.1"
                     /db_xref="GI:10312148"
                     /translation="MIFDYPKPPGLADHLRGRLFGENASRPAEIGTAAAVDEPIAIVA
                     MACRFPGGVHSPEDLWRLVADGVDAVTEFPATRGWDTDRLYHEDPDHEGTTYVRHGAF
```

LDDAGGFDAAFFGISPNEALAMDPQQRLLLETSGKLFERAAIDPTTLAGQDIGVFAGV
NSHDYSMRTHGAAGVEGFRLTGASASVLSGRVAYHFGVEGPAVTVDTACSSSLVALHM
AVQALQRGECAMALAGGVMVMGTVETFVEFSRQRGLAPDGRCKAFADGADGTGWSEGV
GLLLVERLSEARRRGHQVLAVVRGSAVNSDGASNGLTAPNGPAQQRVIRKALAAAGLS
PSDVDAVEAHGTGTTLGDPIEAEALIATYGQDREQPLWLGSVKSNIGHTQAAAGVAGV
IKMVMAMRHGVLPRTLHVDRPSSHVDWSAGAVELLTEARDWTSDGHPRRAGVSSFGIG
GTNAHVVLEEVPAPVATPRPEPAGFLVPVLLSARSAAGLRGQAARLAGFLGAEVRVPD
VAHALATTRAQLDHRAVVLASDRDRLRADLAAFGPGVVTGTPVDGKLAVLFTGQGSQW
AGMGRELAGTFPVFRNAFEAACHAVDAHLGRAPVREVVFDDRTRLDQTMYTQGALFAV
ETALFRLFASWGVRPDLLAGHSIGELAAAHVSGVLDLAEAGELVAARGRLMQALPAGG
AMVAVQAAEDEVAPLLDGTVCVAAVNGPDSVVLSGAEDAVLAVAAGLAARGRKTKRLP
VSHAFHSPLMEPMLGDFRAVAERLTYRAGTLPVVSTLTGKLAALDTPGYWVRQVRDAV
RFGDAVTELNAQGATTYLELGPGGALAAMALGTLGGPEQSCVATLRKNGAEVPGVLTA
LAELHVRGVTVDWTALLDRPGTAVGTVLPTYAFQHERYWVGTDETPSPASPVAARPAE
PVHDVPALVRESVAVVLGHRDTAAPDLDRSFKDLGFDSLGAIKLRNRLRDFTGVDLPS
TLIFDYPNAAVLADHLRAELRGERPAVAAPVVRDVSDEPIAIVGMSTRLPGGADSPEE
LWNLVAEGRDAISGFPVDRGWDLDGLYHPDPAHPGTSYTRTGGFLHDAAQFDAGLFGI
SPREALAMDPQQRLLLETSWEALERAGVDPLSARGSDVGVFTGIVHHDYVTRLREVPE
DVQGYTMTGTASSVASGRVAYVFGFEGPAVTVDTACSSSLVAIHLAAQALRQGECSMA
LAGGATVMASPDAFLEFSRQRGLSADGRCKAYAEGADGTGWAEGVGVVVLERLSVARE
RGH"

BASE COUNT 475 a 1474 c 1437 g 514 t
ORIGIN

Query Match 54.9%; Score 55.4; DB 1; Length 3900;
Best Local Similarity 73.2%; Pred. No. 0.52;
Matches 71; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy       5 GGTGTCGATCGCGGCGGTCAATGGGCCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGC 64
           |||||   |||| |||||||| || |||||   ||||||| || ||||||  ||  | ||
Db    2208 GGTGTGCGTCGCCGCGGTCAACGGCCCGGACTCGGTGGTGCTCTCGGGCGCCGAAGACGC 2267

Qy      65 GGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGC 101
            ||||  ||  ||||||| ||  | ||||| ||||||
Db    2268 CGTGCTCGCCGTCGCGGCCGGACTGGCGGCCCGCGGC 2304
```

Search completed: November 5, 2002, 12:18:55
Job time : 592.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450
Perfect score: 101
Sequence: 1 gatgcacgcgctttcgcccg.....cgtggtggtgctcaagcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	43.4	43.0	516	9	AI532314	AI532314 SD03777.5
	2	37.2	36.8	470	12	BF429029	BF429029 WHE1704_C
	3	36.6	36.2	531	14	BQ245189	BQ245189 TaE15029D
	4	35.6	35.2	597	13	BM488815	BM488815 pgm2n.pk0
c	5	34.6	34.3	885	17	AG159162	AG159162 Pan trogl
	6	34.2	33.9	932	17	CNS0072Q	AL066742 Drosophil
	7	34.2	33.9	934	14	BQ674384	BQ674384 AGENCOURT
	8	34	33.7	715	14	BQ805311	BQ805311 WHE3565_D
c	9	33.8	33.5	284	17	AQ797196	AQ797196 RPCI-21-5
c	10	33.8	33.5	310	17	BH173199	BH173199 0011A14 m
	11	33.8	33.5	640	13	BI869677	BI869677 603393448
	12	33.8	33.5	943	12	BG118354	BG118354 602350162
c	13	33.6	33.3	1580	17	AG083426	AG083426 Pan trogl
c	14	33.4	33.1	232	14	BQ841294	BQ841294 WHE4209_C
c	15	33.4	33.1	1276	17	AG073804	AG073804 Pan trogl
	16	33.2	32.9	1152	14	BM925325	BM925325 AGENCOURT
	17	33	32.7	171	17	AZ916066	AZ916066 PstI_2_d6
c	18	33	32.7	688	17	BH838383	BH838383 LMCR10001
	19	33	32.7	960	17	CNS038VR	AL233136 Tetraodon
	20	33	32.7	1037	17	AG061836	AG061836 Pan trogl
	21	32.8	32.5	1188	14	BM908892	BM908892 AGENCOURT
c	22	32.8	32.5	1206	14	BQ956542	BQ956542 AGENCOURT
	23	32.6	32.3	422	9	AL505522	AL505522 AL505522
	24	32.6	32.3	513	12	BG753881	BG753881 602732802
c	25	32.6	32.3	553	10	AW285052	AW285052 LG1_297_G
	26	32.6	32.3	1094	17	AG099072	AG099072 Pan trogl
c	27	32.4	32.1	521	12	BF628325	BF628325 HVSMEb000
	28	32.4	32.1	959	17	AG065673	AG065673 Pan trogl
c	29	32.4	32.1	1096	14	BM910915	BM910915 AGENCOURT
c	30	32.2	31.9	155	14	BQ468858	BQ468858 HM02J05r
c	31	32.2	31.9	818	12	BF620794	BF620794 HVSMEc002
c	32	32.2	31.9	1615	14	BM911635	BM911635 AGENCOURT
	33	32	31.7	343	10	BE426531	BE426531 WHE0318_E
c	34	32	31.7	457	10	BE125336	BE125336 DG1_19_E0
	35	32	31.7	460	14	BQ346027	BQ346027 PM3-NT031
	36	32	31.7	505	13	BI776710	BI776710 EBpi07_SQ
	37	32	31.7	507	10	BE419271	BE419271 WWR06.B1R
	38	32	31.7	626	13	BM389954	BM389954 UI-R-CN1-
	39	32	31.7	787	9	AU134757	AU134757 AU134757
	40	32	31.7	821	12	BF620615	BF620615 HVSMEc002
	41	32	31.7	822	12	BG299905	BG299905 HVSMEa002
c	42	32	31.7	844	17	AG063094	AG063094 Pan trogl
	43	32	31.7	850	13	BI258402	BI258402 602972664
	44	32	31.7	1024	13	BM550188	BM550188 AGENCOURT
c	45	32	31.7	1057	17	CNS046BO	AL276477 Tetraodon

ALIGNMENTS

RESULT 1
AI532314
LOCUS AI532314 516 bp mRNA linear EST 19-APR-2001
DEFINITION SD03777.5prime SD Drosophila melanogaster Schneider L2 cell culture pOT2 Drosophila melanogaster cDNA clone SD03777 5prime, mRNA sequence.
ACCESSION AI532314
VERSION AI532314.1 GI:4446449
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 516)
AUTHORS Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.
TITLE BDGP/HHMI Drosophila EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
hit genomic sequence AC005554
Plate: 37 row: G column: 5
High quality sequence stop: 514.
FEATURES Location/Qualifiers

```
     source          1. .516
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="SD03777"
                     /clone_lib="SD Drosophila melanogaster Schneider L2 cell
                     culture pOT2"
                     /lab_host="DH5-alpha"
                     /note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
                     fractionated cDNAs were directly ligated into pOT2.
                     Plasmid cDNA library."
BASE COUNT      101 a    163 c    158 g     94 t
ORIGIN

  Query Match           43.0%;  Score 43.4;  DB 9;  Length 516;
  Best Local Similarity   69.4%;  Pred. No. 1.5;
  Matches   59;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      17 CCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGCCGAGGGC 76
           | || |||    |||||||| |||   || |  ||  ||||||| || ||||| ||   |
Db     201 CACGACGGCTCCTGCAAGACCTTCGATGCGGCGGCCAACGGCTATGCCCGGGCGGACACC 260

Qy      77 TGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||| ||| ||||  || |||
Db     261 TGCGCCGTTGTGCTGCTTCAGAGGC 285


RESULT 2
BF429029
LOCUS       BF429029                 470 bp    mRNA    linear   EST 29-NOV-2000
DEFINITION  WHE1704_C03_F06ZS Wheat heat stressed spike cDNA library Triticum
            aestivum cDNA clone WHE1704_C03_F06, mRNA sequence.
ACCESSION   BF429029
VERSION     BF429029.1  GI:11440948
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 470)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Malatrasi,M., Miller,R.,
            Nguyen,H.T., Rausch,C.J., Seaton,C.L., Tong,J.C. and Zhang,D.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Heat stressed spike cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture. Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .470
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE1704_C03_F06"
                     /clone_lib="Wheat heat stressed spike cDNA library"
                     /tissue_type="Whole spike"
                     /dev_stage="Spikes at 5, 10, 15 and 20 days after
                     anthesis"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Spikes at 5, 10, 15 and 20
                     days after anthesis were heat stressed under  two
                     conditions at Texas Tech University (D. Zhang in HT Nguyrn
                     lab):  (1) at 38 C for 4 hours and (2) 5 days of cyclic
                     treatment of 38 C  for 4 hours.  Total RNA and poly(A) RNA
                     were prepared, a cDNA library  was made, and the cDNA
                     clones were in vivo excised to give pBluescript  phagemids
                     in the TJ Close lab (Choi, Close, Fenton, Malatrasi) at
                     the  University of California, Riverside.  Plasmid DNA
                     preparations and  DNA sequencing were performed in the OD
                     Anderson lab (all other authors)."
BASE COUNT       46 a    220 c    138 g     65 t      1 others
ORIGIN

  Query Match           36.8%;  Score 37.2;  DB 12;  Length 470;
  Best Local Similarity   62.6%;  Pred. No. 35;
  Matches   57;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           | | |||||   | ||    ||||| || ||| |  | || ||  ||||||   || ||
Db     346 GCTCCACGCCGCTGCGGTGCGCGGGGGGAGCANGTGGCTCGCGAGCGACGCCTCCGCCTC 405

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTG 91
           ||||  |   |||  || ||||| | |||||
Db     406 CGCGGCGAGGGAGCCCTCCGCCGAGCTGGTG 436


RESULT 3
BQ245189
LOCUS       BQ245189                 531 bp    mRNA    linear   EST 03-MAY-2002
DEFINITION  TaE15029D12R TaE15 Triticum aestivum cDNA clone TaE15029D12R, mRNA
            sequence.
ACCESSION   BQ245189
VERSION     BQ245189.1  GI:20441065
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 531)
  AUTHORS   Cloutier,S.
  TITLE     Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average insert size is >1.4 kb
            Plate: 029  row: D  column: 12
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .531
                     /organism="Triticum aestivum"
                     /cultivar="Glenlea"
                     /db_xref="taxon:4565"
                     /clone="TaE15029D12R"
                     /clone_lib="TaE15"
                     /tissue_type="developing seeds"
                     /dev_stage="15 days after anthesis"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies);
                     Site_1: NotI; Site_2: MluI; mRNA obtained from wheat seeds
                     of cultivar Glenlea 15 days post-anthesis"
BASE COUNT      110 a    151 c    164 g    106 t
ORIGIN

  Query Match           36.2%;  Score 36.6;  DB 14;  Length 531;
  Best Local Similarity   62.6%;  Pred. No. 47;
  Matches   57;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           | | |||||   | ||    ||||| || ||| |  | || ||  ||||||   || ||
Db      43 GGTCCACGCCGCTGCGGTGCGCGGGGGGAGCAGGTGGCTCGCGAGCGACGCCTCCGCCTC 102
```

```
Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTG 91
           ||||  |   |||  || ||||| | |||||
Db     103 CGCGGCGAGGGAGCCCTCCGCCGAGCTGGTG 133


RESULT 4
BM488815
LOCUS       BM488815                597 bp    mRNA    linear   EST 07-FEB-2002
DEFINITION  pgm2n.pk008.p20 Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA
            clone pgm2n.pk008.p20 5' similar to ref|NP_446119.1 (NM_053667)
            growth supressor 1; leprecan [Rattus norvegicus]
            gb|AAD51875.1|AF087433_1 (AF087433) leprecan [Rattus norvegicus],
            mRNA sequence.
ACCESSION   BM488815
VERSION     BM488815.1  GI:18609746
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Cogburn,L.A. and Monsonego-Ornan,E.
  TITLE     ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
            Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome
            Project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Gallus gallus"
                     /strain="Commercial broiler and  Ottawa Res. Centre
                     Strains 90 & 21"
                     /db_xref="taxon:9031"
                     /clone="pgm2n.pk008.p20"
                     /clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
                     and Epiphyseal Growth Plate cDNA library (pgm2n)"
                     /sex="Male and Female"
                     /tissue_type="Breast muscle, leg muscle and epiphyseal
                     growth plate"
                     /dev_stage="Breast,leg:Embryo(d19);post-hatch(1d,1,3,5,7,9
                     ,11 weeks);growth plate(1d,7d,14d post-hatch)"
                     /lab_host="E. coli EMDH10B"
                     /note="Vector: pCMVSPORT6; Library made from equivalent
                     pools of total RNA isolated from each tissue (embryonic
                     muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
                     plate 33.3% of the final RNA pool). Single pass sequencing
                     from 5'-end"
BASE COUNT       78 a    200 c    199 g     90 t     30 others
ORIGIN

  Query Match             35.2%;  Score 35.6;  DB 13;  Length 597;
  Best Local Similarity   60.2%;  Pred. No. 77;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           || | | ||  || |||     | |  |  |   ||| | |||| ||| || |  |||||
Db      68 GCCCCCTCTGCCGGCCGAGCCCCCGGACGCGCTCTTCGCTGCCGGCGCCGAGGCGTACGC 127

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||  ||  |   |||||||||| |||  | |  ||
Db     128 GCGGGGGGACTGGCCCGCCGTGGTGCTGCAGATGGAGC 165


RESULT 5
AG159162/c
LOCUS       AG159162                885 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-024J17.TJ, genomic survey
            sequence.
ACCESSION   AG159162
VERSION     AG159162.1  GI:16688840
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-024J17.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 885)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                Sequencing: TJ
             LIBRARY
                Vector    : pBACe3.6
                R.Site 1  : EcoRI
                R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .885
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-024J17.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT       14 a    473 c    357 g     16 t     25 others
ORIGIN

  Query Match             34.3%;  Score 34.6;  DB 17;  Length 885;
  Best Local Similarity   62.7%;  Pred. No. 1.2e+02;
  Matches   52;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           || ||||    ||| | ||||| || ||  | ||  |  | |||  || |  ||   ||
Db     565 GCGCGCGGGGGCGCGCCGCGGGGGGCGCGGGGCGCCCGGGNCCGGGGCCGGGGGGCGCGG 506

Qy      64 GCGGGCCGAGGGCTGCGCCGTGG 86
           | | |||| |||| ||||||  ||
Db     505 GGGCGCCGCGGGCGGCGCCGNGG 483


RESULT 6
CNS0072Q
LOCUS       CNS0072Q                932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .932
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="BACR14B09"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
BASE COUNT      155 a    202 c    241 g     91 t    243 others
ORIGIN

  Query Match             33.9%;  Score 34.2;  DB 17;  Length 932;
  Best Local Similarity   46.4%;  Pred. No. 1.5e+02;
  Matches   45;  Conservative   17;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 CACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 64
           | :||:|   ||| ||:|||:| | ||  :     | |: ||: : |   |   | |||:
Db     692 CSSGCSCCCCCGCGCGSCGGSCCGCGCSCSGGCSSCGCSCCCSCSCCCCSCCCCSSCGCS 751

Qy      65 CGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           :::|| | || : |||:|| |: || || :  ||||:
Db     752 SSSGCGGSGGCSSGCGSCGCGSCGGSGCGSCGGCGGS 788


RESULT 7
BQ674384
LOCUS       BQ674384                 934 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8211307 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6276688
            5', mRNA sequence.
ACCESSION   BQ674384
VERSION     BQ674384.1  GI:21785218
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 934)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2461  row: n  column: 17
            High quality sequence stop: 464.
FEATURES              Location/Qualifiers
     source           1. .934
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6276688"
                      /clone_lib="NIH_MGC_102"
                      /tissue_type="epidermoid carcinoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                      Site_2: EcoRI; cDNA made by oligo-dT priming.
                      Directionally cloned into EcoRI/XhoI sites using the
                      following 5' adaptor: GGCACGAG(G).  Library constructed
                      by Ling Hong in the laboratory of Gerald M. Rubin
                      (University of  California, Berkeley) using ZAP-cDNA
                      synthesis kit (Stratagene) and Superscript II RT (Life
                      Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      125 a    361 c    347 g     96 t      5 others
ORIGIN

  Query Match             33.9%;  Score 34.2;  DB 14;  Length 934;
  Best Local Similarity   60.7%;  Pred. No. 1.5e+02;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       3 TGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACG 62
           ||| ||  | |||||||| |||| |  ||  |  || |   || |  | |  || |   |
Db     526 TGCCCGGCCCTTCGCCCGCCGGGGGCCGCCCGCGGTGCCGCGCGGGGGNGCCCGCCCCGG 585

Qy      63 CGCGGGCCGAGGGCTGCGCCGTGGTGGTG 91
           | |||| ||  ||| |||  ||| ||| |
Db     586 CCCGGGGCGCCGGCGGCGGGGTGNTGGGG 614


RESULT 8
BQ805311
LOCUS       BQ805311                 715 bp    mRNA    linear   EST 31-JUL-2002
DEFINITION  WHE3565_D05_H09ZS Wheat developing grains cDNA library Triticum
            aestivum cDNA clone WHE3565_D05_H09, mRNA sequence.
ACCESSION   BQ805311
VERSION     BQ805311.1  GI:22029520
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Altenbach,S., Anderson,O.D., Chao,S., Chin,A., Close,T.J., Cronin
            ,K., Crossman,C., Fenton,R.D., Lazo,G.R., Pham,J., Rausch,C.J.,
            Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Developing grains cDNA library
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES              Location/Qualifiers
     source           1. .715
                      /organism="Triticum aestivum"
                      /cultivar="Butte 86"
                      /db_xref="taxon:4565"
                      /clone="WHE3565_D05_H09"
                      /clone_lib="Wheat developing grains cDNA library"
                      /tissue_type="whole grains"
                      /dev_stage="3-44 days post anthesis seed"
                      /lab_host="E. coli SOLR"
                      /note="Vector: Lambda ZAP II, excised phagemid; Site_1:
                      EcoRI; Plants were grown under six following different
                      environmental regimes in greenhouse, Environment 1)
                      24oC/17oC day/night, well-watered, with post-anthesis
```

fertilizer, Environment 2) 24oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 3) 37oC/17oC day/night, well-watered, with post-anthesis fertilizer, Environment 4) 37oC/17oC day/night, well-watered, without post-anthesis fertilizer, Environment 5) 37oC/17oC day/night plus drought, with post-anthesis fertilizer, Environment 6) 37oC/17oC day/night plus drought, without post-anthesis fertilizer, developing wheat grains from the following were excised and frozen in liquid nitrogen, Environment 1 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 2 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 36, 40, 44 DPA Environment 3 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 4 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 32, 34 DPA Environment 5 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA Environment 6 at 3, 5, 7, 8, 10, 12, 16, 20, 24, 28, 30 DPA and total RNA was prepared by S. Altenbach and K. Cronin at USDA-ARS, Albany, CA. A cDNA library was made using poly (A) RNA, and the cDNA clones were in vivo excised to give pBluescript SK(-) phagemids in the TJ Close lab (Chin, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (others)."

BASE COUNT 135 a 216 c 267 g 97 t
ORIGIN

Query Match 33.7%; Score 34; DB 14; Length 715;
Best Local Similarity 63.4%; Pred. No. 1.7e+02;
Matches 52; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy      20 GGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGCCGAGGGCTGC 79
           |||||  |  || ||| |    |||||||||  || || |||  |||||||||| || | |
Db     180 GGCGGCGGCGGCGAGATGAGGGCGGCCGACGGCGAGGGGTACCTGCGGGCCGACGGATTC 239

Qy      80 GCCGTGGTGGTGCTCAAGCGGC 101
           | | | ||    ||| |   ||
Db     240 GACCTCGTCAACCTCGACATGC 261
```

RESULT 9
AQ797196/c
LOCUS AQ797196 284 bp DNA linear GSS 26-JAN-2001
DEFINITION RPCI-21-521I14-T7 RPCI-21 Mouse (129S6/SvEvTac) genomic library Mus musculus genomic clone RPCI-21-521I14-T7 similar to Catalase (CAT) gene : Mouse Chr 2.57, DNA sequence.
ACCESSION AQ797196
VERSION AQ797196.1 GI:5705665
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 284)
AUTHORS Osoegawa,K., Tateno,M., Woon,P.Y., Frengen,E., Mammoser,A.G., Catanese,J.J., Hayashizaki,Y. and de Jong,P.J.
TITLE Bacterial artificial chromosome libraries for mouse sequencing and functional analysis
JOURNAL Genome Res. 10 (1), 116-128 (2000)
MEDLINE 20113121
COMMENT Contact: de Jong, P.J.
Children's Hospital Oakland Research Institute
747 Fifty second Street, Oakland, CA 94609-1809, USA
Tel: 510 450 7911
Fax: 510 450 7924
Email: pdejong@mail.cho.org
PAC end sequences. For clone availability please contact Pieter de Jong (pdejong@.mail.cho.org). BACPAC Resources WWW site: www.chori.org/bacpac
Seq primer: T7
Class: PAC end.
FEATURES Location/Qualifiers
source 1. .284
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RPCI-21-521I14-T7"
/clone_lib="RPCI-21 Mouse (129S6/SvEvTac) genomic library"
/sex="Female"
/cell_type="Spleen"
/note="Vector: pPAC4; PAC clones in E. coli DH10B"
BASE COUNT 57 a 100 c 85 g 42 t
ORIGIN

Query Match 33.5%; Score 33.8; DB 17; Length 284;
Best Local Similarity 62.4%; Pred. No. 2.2e+02;
Matches 53; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy      13 TTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGCCGA 72
           | | ||||    | ||||||    |   |  |||   | | || |||  |||| ||||||
Db     266 TGCACCCGCTTCGAGGTGCAGAGTGCGGTGCGCCTTGGTGCACAGCTGGGCGCTGGCCGA 207

Qy      73 GGGCTGCGCCGTGGTGGTGCTCAAG 97
           | |||||||  |||  | | |||||
Db     206 GAGCTGCGCACTGGCAGCGGTCAAG 182
```

RESULT 10
BH173199/c
LOCUS BH173199 310 bp DNA linear GSS 09-OCT-2001
DEFINITION 0011A14 maize genomic ORF library A Zea mays genomic, DNA sequence.
ACCESSION BH173199
VERSION BH173199.1 GI:15988563
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 310)
AUTHORS Fu,Y., Liu,F. and Schnable,F.S.
TITLE Maize genomic sequences rescued by a novel expression vector system
JOURNAL Unpublished (2001)
COMMENT Contact: Patrick S. Schnable
Iowa State University
G405 Agronomy, Ames, IA 50011, USA
Tel: 515 294 0975
Fax: 515 294 2299
Email: schnable@iastate.edu
Size-selected, PstI-digested DNA (100-800 bp) from immature, unpollinated ears was cloned into the PstI site of pHis4.
Class: plasmid ends.
FEATURES Location/Qualifiers
source 1. .310
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_lib="maize genomic ORF library A"
/note="Vector: pHis4; Site_1: PstI; Site_2: SmaI"
BASE COUNT 23 a 153 c 96 g 38 t
ORIGIN

Query Match 33.5%; Score 33.8; DB 17; Length 310;
Best Local Similarity 58.4%; Pred. No. 2.1e+02;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           | ||  |||||    | |||||| |||||  | ||||    ||||||||||| |
Db     239 GGTGAGCGCGCCCGTGGCCGGCGCGCGGTAGAGGACGAGGCCGGCCGACGCGGCCACGGG 180

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           | ||   |||   |||   ||   |  |  | || ||| ||
Db     179 CACGGCAGCCCCCGGCGCGGCGTGGCAGCGGTTCCAGCTGC 139
```

RESULT 11
BI869677
LOCUS BI869677 640 bp mRNA linear EST 11-OCT-2001

```
DEFINITION  603393448F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5403514 5',
            mRNA sequence.
ACCESSION   BI869677
VERSION     BI869677.1  GI:16043350
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 640)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12028  row: p  column: 11
            High quality sequence start: 6
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .640
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5403514"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      140 a    191 c    202 g    107 t
ORIGIN

  Query Match             33.5%;  Score 33.8;  DB 13;  Length 640;
  Best Local Similarity   64.9%;  Pred. No. 1.9e+02;
  Matches   50;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           |||||| |   | |  |||||||||| || ||| | ||  |  |  ||  || | || |||
Db     153 GATGCAGGGCTTGTACCCCGGCGGGGGGGGCATGGCGGGCCAGAGCGCGGCCGGCGTCTA 212

Qy      61 CGCGCGGGCCGAGGGCT 77
           ||||  |||   |||||
Db     213 CGCGGCGGCTATGGGCT 229


RESULT 12
BG118354
LOCUS       BG118354                 943 bp    mRNA    linear   EST 30-JAN-2001
DEFINITION  602350162F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4444958 5',
            mRNA sequence.
ACCESSION   BG118354
VERSION     BG118354.1  GI:12611860
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 943)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10221  row: l  column: 15
            High quality sequence stop: 625.
FEATURES             Location/Qualifiers
     source          1. .943
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4444958"
                     /clone_lib="NIH_MGC_90"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.7 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      227 a    269 c    273 g    174 t
ORIGIN

  Query Match             33.5%;  Score 33.8;  DB 12;  Length 943;
  Best Local Similarity   64.9%;  Pred. No. 1.8e+02;
  Matches   50;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           |||||| |   | |  |||||||||| || ||| | ||  |  |  ||  || | || |||
Db     268 GATGCAGGGCTTGTACCCCGGCGGGGGGGGCATGGCGGGCCAGAGCGCGGCCGGCGTCTA 327

Qy      61 CGCGCGGGCCGAGGGCT 77
           ||||  |||   |||||
Db     328 CGCGGCGGCTATGGGCT 344


RESULT 13
AG083426/c
LOCUS       AG083426                 1580 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-080N21.R, genomic survey sequence.
ACCESSION   AG083426
VERSION     AG083426.1  GI:16635228
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-080N21.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1580)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
```

```
     source          1. .1580
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-080N21.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       277 a    584 c    609 g     55 t     55 others
ORIGIN

  Query Match             33.3%;  Score 33.6;  DB 17;  Length 1580;
  Best Local Similarity   63.2%;  Pred. No. 1.8e+02;
  Matches   48;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||    | ||| |||| || ||  | || |  ||||   ||||| |     ||||| |
Db     296 GCGCGCCGGNCCGNCGGGGGGCGCGGGGCGCTGCCGGCGCCCGCGGCCNNGGGCGCGCCG 237

Qy      68 GCCGAGGGCTGCGCCG 83
               ||| |||  ||||||
Db     236 CCCGCGGGGCGCGCCG 221


RESULT 14
BQ841294/c
LOCUS       BQ841294                 232 bp    mRNA    linear   EST 12-AUG-2002
DEFINITION  WHE4209_C11_E21ZS Aegilops speltoides pre-meiotic anther cDNA
            library Aegilops speltoides cDNA clone WHE4209_C11_E21, mRNA
            sequence.
ACCESSION   BQ841294
VERSION     BQ841294.1  GI:22210703
KEYWORDS    EST.
SOURCE      Aegilops speltoides.
  ORGANISM  Aegilops speltoides
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Aegilops.
REFERENCE   1  (bases 1 to 232)
  AUTHORS   Akhunov,E., Anderson,O.D., Chao,S., Crossman,C., Devorak,J., Lazo
            ,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Pre-meiotic anther cDNA library from Aegilops speltoides
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .232
                     /organism="Aegilops speltoides"
                     /cultivar="F2 from 2-12-4-8-1-1-1-(1) x PI36909-12-811-(1
                     )"
                     /db_xref="taxon:4573"
                     /clone="WHE4209_C11_E21"
                     /clone_lib="Aegilops speltoides pre-meiotic anther cDNA
                     library"
                     /tissue_type="Anther"
                     /dev_stage="Pre-meiotic anthers"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in a growth
                     chamber at the University of California, Davis (Akhunov in
                     Devorak's lab). Pre-meiotic anthers were harvested, total
                     RNA and poly(A) RNA were prepared, a cDNA library was made
                     , and the cDNA clones were in vivo excised to give
                     pBluescript phagemids at the University of California,
                     Davis (Akhunov in Devorak's lab). Plasmid DNA preparations
                     and DNA sequencing were performed in the OD Anderson lab
                     (all other authors)."
BASE COUNT       38 a     95 c     79 g     20 t
ORIGIN

  Query Match             33.1%;  Score 33.4;  DB 14;  Length 232;
  Best Local Similarity   58.6%;  Pred. No. 2.7e+02;
  Matches   58;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       3 TGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACG 62
           ||| | || | ||| |||| |||| || |  | | | |||||| ||| |
Db     221 TGCGCACGGTGTCGACCGGGTCGCGGCGCGCCGGCCGCTTGCAGCAGGCGGACTGCTGCT 162

Qy      63 CGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             |  |||  | || ||||||  |||  ||  | |||||
Db     161 GCCACGCCCGGCGCCGCGCCGCCGTGACGCGGAGGCGGC 123


RESULT 15
AG073804/c
LOCUS       AG073804                 1276 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-065L10.R, genomic survey sequence.
ACCESSION   AG073804
VERSION     AG073804.1  GI:16625606
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-065L10.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1276)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1276
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-065L10.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       92 a    440 c    376 g    258 t    110 others
ORIGIN

  Query Match             33.1%;  Score 33.4;  DB 17;  Length 1276;
  Best Local Similarity   53.8%;  Pred. No. 2.1e+02;
  Matches   49;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           || |||||   ||| |||||| ||| ||  |  |  | |||| | ||||| |  |  | |
Db     433 GCGCGCGCGCGCGCTCGGCGGNCGGGGCTNGGNGCNCGCGGCGGCCGCGGCCNCCCTCNC 374

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTC 94
```

```
Db    373 NCCCGNCCCGCCCGNNCCGNNGGGGNNNCNC 343
```

Search completed: November 5, 2002, 16:17:53
Job time : 1120.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450
Perfect score: 101
Sequence: 1 gatgcacgcgctttcgcccg.....cgtggtggtgctcaagcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	56.4	55.8	15872	10	US-09-861-289-1	Sequence 1, Appli
	2	46.8	46.3	13842	10	US-09-861-289-30	Sequence 30, Appl
	3	46.8	46.3	36778	10	US-09-861-289-5	Sequence 5, Appli
	4	43.6	43.2	4041	10	US-09-861-289-36	Sequence 36, Appl
	5	43.6	43.2	11220	10	US-09-861-289-32	Sequence 32, Appl
	6	42	41.6	4689	10	US-09-861-289-34	Sequence 34, Appl
	7	41.8	41.4	658	10	US-09-924-256A-69	Sequence 69, Appl
	8	38.8	38.4	643	10	US-09-924-256A-53	Sequence 53, Appl
	9	32.4	32.1	643	10	US-09-924-256A-29	Sequence 29, Appl
	10	32	31.7	3966	10	US-09-764-855-227	Sequence 227, App
	11	30.8	30.5	1281	10	US-09-861-289-19	Sequence 19, Appl
	12	30.8	30.5	13613	10	US-09-861-289-3	Sequence 3, Appli
	13	30.4	30.1	753	10	US-09-805-177-1	Sequence 1, Appli
	14	30.4	30.1	761	10	US-09-924-256A-79	Sequence 79, Appl
	15	30.4	30.1	1359	10	US-09-954-456-694	Sequence 694, App
	16	30	29.7	765	10	US-09-924-256A-57	Sequence 57, Appl
c	17	30	29.7	817	12	US-10-062-254-217	Sequence 217, App
	18	29.8	29.5	655	10	US-09-924-256A-45	Sequence 45, Appl
	19	29.8	29.5	722	10	US-09-924-256A-49	Sequence 49, Appl
	20	29.4	29.1	703	10	US-09-924-256A-51	Sequence 51, Appl
	21	29.2	28.9	1152	10	US-09-147-346-1	Sequence 1, Appli
	22	29.2	28.9	2760	10	US-09-452-599-20	Sequence 20, Appl
	23	29	28.7	381	10	US-09-878-574-1137	Sequence 1137, Ap
c	24	29	28.7	3571	10	US-09-823-033-3	Sequence 3, Appli
	25	28.8	28.5	439	10	US-09-764-877-2558	Sequence 2558, Ap
	26	28.8	28.5	439	10	US-09-764-877-2559	Sequence 2559, Ap
	27	28.8	28.5	927	10	US-09-815-242-7757	Sequence 7757, Ap
	28	28.8	28.5	1277	10	US-09-880-107-2230	Sequence 2230, Ap
c	29	28.8	28.5	1383	10	US-09-741-669-271	Sequence 271, App
	30	28.4	28.1	404	10	US-09-925-299-394	Sequence 394, App
	31	28.4	28.1	435	10	US-09-960-352-175	Sequence 175, App
c	32	28.4	28.1	933	9	US-09-934-392-3	Sequence 3, Appli
c	33	28.4	28.1	1543	9	US-09-934-392-1	Sequence 1, Appli
	34	28.4	28.1	1566	10	US-09-833-381-987	Sequence 987, App
	35	28.2	27.9	643	10	US-09-924-256A-31	Sequence 31, Appl
	36	28	27.7	643	10	US-09-924-256A-43	Sequence 43, Appl
c	37	28	27.7	4041	10	US-09-861-289-36	Sequence 36, Appl
c	38	28	27.7	36778	10	US-09-861-289-5	Sequence 5, Appli
	39	27.8	27.5	2126	12	US-10-052-586-495	Sequence 495, App
	40	27.8	27.5	2261	10	US-09-823-356-21	Sequence 21, Appl
c	41	27.6	27.3	888	10	US-09-815-242-7873	Sequence 7873, Ap
	42	27.6	27.3	5184	10	US-09-845-583-9	Sequence 9, Appli
c	43	27.4	27.1	391	10	US-09-764-847-344	Sequence 344, App
	44	27.4	27.1	1656	12	US-10-033-109-13	Sequence 13, Appl
	45	27.4	27.1	2087	10	US-09-349-385-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 55.8%; Score 56.4; DB 10; Length 15872;
Best Local Similarity 73.5%; Pred. No. 5.7e-07;
Matches 72; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           | | | |||  |  ||| ||| || |||||| |||||||||||| ||| ||||||  ||
Db    9142 GAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGG 9201

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             |||||||||||  || |||| | || ||  ||||||
Db    9202 CTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGC 9239
```

RESULT 2
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             46.3%;  Score 46.8;  DB 10;  Length 13842;
  Best Local Similarity   67.3%;  Pred. No. 0.00017;
  Matches   66;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           ||  |||||  | || | ||| ||  ||||| | |||||||| ||||| ||||| | ||
Db    8454 GCGGGCGCTCGCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGG 8513

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             || ||||||  || | || || |  |  ||||||
Db    8514 CGCCGCGGAGGGCGTCGGCCTGCTGCTCGTGGAGCGGC 8551


RESULT 3
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             46.3%;  Score 46.8;  DB 10;  Length 36778;
  Best Local Similarity   67.3%;  Pred. No. 0.00016;
  Matches   66;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           ||  |||||  | || | ||| ||  ||||| | |||||||| ||||| ||||| | ||
Db   10195 GCGGGCGCTCGCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGG 10254

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             || ||||||  || | || || |  |  ||||||
Db   10255 CGCCGCGGAGGGCGTCGGCCTGCTGCTCGTGGAGCGGC 10292


RESULT 4
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-36

  Query Match             43.2%;  Score 43.6;  DB 10;  Length 4041;
  Best Local Similarity   65.3%;  Pred. No. 0.0012;
  Matches   64;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           ||| ||  |  |  ||| ||| |||  ||||   ||| |  | |  |||||||||  |||
Db     747 GCAGGCCATGGCCGCCGACGGCCGGACCAAGGGCTTCGCCTCGGCGGCGGACGGCCTCGC 806

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             ||| |||||||  ||||||  |  | ||| | ||||
Db     807 CTGGGGCGAGGGCGTCGCCGTACTCCTCCTCGAACGGC 844


RESULT 5
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             43.2%;  Score 43.6;  DB 10;  Length 11220;
  Best Local Similarity   65.3%;  Pred. No. 0.0012;
  Matches   64;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           || || |||  ||   | ||| ||||  ||| ||||| | ||    |||||||||| ||
Db    5262 GCGCGGGCTCGCGGAGGACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCTTCGG 5321

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             | ||| ||||||  || | || |  |  || |||| |
Db    5322 CCCGGCGGAGGGCGTCGGCATGCTCCTCGTCGAGCGCC 5359


RESULT 6
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
```

```
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             41.6%;  Score 42;  DB 10;  Length 4689;
  Best Local Similarity   64.3%;  Pred. No. 0.0032;
  Matches   63;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           || || |||  |  | | ||| ||||  ||| |||||  |  ||   |||||||||| ||
Db     753 GCGCGGGCTGGCCGCGGACGGCCGGTCGAAGGCGTTCGCCACCTCGGCGGACGGCTTCGG 812

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             |  || ||||||  ||  ||  || || |  |||| |
Db     813 CCCCGCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCC 850


RESULT 7
US-09-924-256A-69
; Sequence 69, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
;   LENGTH: 658
;   TYPE: DNA
;   ORGANISM: Usnea florida
US-09-924-256A-69

  Query Match             41.4%;  Score 41.8;  DB 10;  Length 658;
  Best Local Similarity   63.4%;  Pred. No. 0.004;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           | || |   |||||| ||    || || | |   | ||     || |||||||| |||||
Db     367 GCTGAAAATGCTTTCTCCTACGGGCCGCTCCCGCATGTGGGATGCGGACGCGGATGGCTA 426

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           ||| || |  ||||||   || |  || ||||| || ||||
Db     427 CGCTCGTGGGGAGGGCGTAGCGGCTGTAGTGCTGAAACGGC 467


RESULT 8
US-09-924-256A-53
; Sequence 53, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-09-924-256A-53

  Query Match             38.4%;  Score 38.8;  DB 10;  Length 643;
  Best Local Similarity   67.1%;  Pred. No. 0.024;
  Matches   55;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      18 CCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGCCGAGGGCT 77
           | |  || || ||||| || ||    || || || || || |||| ||||| ||| || |
Db     384 CTGATGGTCGTTGCAAAACCTTTGATGCTGATGCTGATGGTTACGTGCGGGGCGAAGGTT 443

Qy      78 GCGCCGTGGTGGTGCTCAAGCG 99
           | |  || ||  | ||||||||
Db     444 GTGGTGTTGTAATTCTCAAGCG 465


RESULT 9
US-09-924-256A-29
; Sequence 29, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Clone
;   OTHER INFORMATION: Xa.cum.6H
US-09-924-256A-29

  Query Match             32.1%;  Score 32.4;  DB 10;  Length 643;
  Best Local Similarity   62.2%;  Pred. No. 1.1;
  Matches   51;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      17 CCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGCCGAGGGC 76
           ||  ||||||  |||||| ||||    ||    || || ||||||   || |  || ||
Db     383 CCTAGCGGGCAATGCAAGCCGTTTGATGCTTCTGCTGATGGCTACTGCCGTGGAGAAGGA 442

Qy      77 TGCGCCGTGGTGGTGCTCAAGC 98
             |  || || ||||| ||||
Db     443 GTTGGTGTCGTTGTGCTTAAGC 464


RESULT 10
US-09-764-855-227
; Sequence 227, Application US/09764855
; Patent No. US20020119919A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA110
;  CURRENT APPLICATION NUMBER: US/09/764,855
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 227
;   LENGTH: 3966
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-855-227

  Query Match             31.7%;  Score 32;  DB 10;  Length 3966;
  Best Local Similarity   58.3%;  Pred  No. 1.2;
  Matches   56;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       2 ATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTAC 61
           | | |||||||   || | ||| | |  |||  ||    |  |||||| | ||||  |  |
Db    3802 AGGAACGCGCTCCGGCTCAGCGTGGGCCCCAGCACCACCAGGGCCGAGGTGGACCTCGTC 3861

Qy      62 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAG 97
           | || || |  |       ||||||| |  |||  ||
Db    3862 GTGCAGGACCTGAAGCAGGCCGTGGCGCAGCTGGAG 3897


RESULT 11
US-09-861-289-19
; Sequence 19, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 1281
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-19

  Query Match             30.5%;  Score 30.8;  DB 10;  Length 1281;
  Best Local Similarity   57.1%;  Pred. No. 2.7;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||||||   |||  || |   |   |  |     |   | ||  |||| |  |||
Db     924 GATGCACGCGCTCCTGCCGAGCTGCTCGGCGATCATCCACCACGGCGGGGCGGGCACCTA 983

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGC 98
           ||||   |||| |  |  ||| |||  |  | ||| ||
Db     984 CGCGACCGCCGTGATCAACGCGGTGCCGCAGGTCATGC 1021


RESULT 12
US-09-861-289-3
; Sequence 3, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 13613
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-3

  Query Match             30.5%;  Score 30.8;  DB 10;  Length 13613;
  Best Local Similarity   57.1%;  Pred. No. 2.3;
  Matches   56;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||||||   |||  || |   |   |  |     |   | ||  |||| |  |||
Db    3085 GATGCACGCGCTCCTGCCGAGCTGCTCGGCGATCATCCACCACGGCGGGGCGGGCACCTA 3144

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGC 98
           ||||   |||| |  |  ||| |||  |  | ||| ||
Db    3145 CGCGACCGCCGTGATCAACGCGGTGCCGCAGGTCATGC 3182


RESULT 13
US-09-805-177-1
; Sequence 1, Application US/09805177
; Patent No. US20020001805A1
; GENERAL INFORMATION:
;  APPLICANT: Richard Bruce Roden
;  APPLICANT:  Honami Naora
;  TITLE OF INVENTION: IMMUNOGENIC OVARIAN CANCER GENES
;  FILE REFERENCE: 031787.0090
;  CURRENT APPLICATION NUMBER: US/09/805,177
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: 60/189,226
;  PRIOR FILING DATE: 2000-03-14
;  PRIOR APPLICATION NUMBER: 60/258,452
;  PRIOR FILING DATE: 2000-12-28
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(654)
US-09-805-177-1

  Query Match             30.1%;  Score 30.4;  DB 10;  Length 753;
  Best Local Similarity   61.2%;  Pred. No. 3.5;
  Matches   49;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||  |   | |  ||||||||| || ||| | ||  |  |  ||  || | || |||
Db     162 GATGCAGGGCTTGTACCCCGGCGGGGGGGGCATGGCGGGCCAGAGCGCGGCCGGCGTCTA 221

Qy      61 CGCGCGGGCCGAGGGCTGCG 80
           ||||   | | | ||   ||
Db     222 CGCGGCCGGCTATGGGCTCG 241


RESULT 14
US-09-924-256A-79
; Sequence 79, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
```

```
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
;   LENGTH: 761
;   TYPE: DNA
;   ORGANISM: Usnea florida
US-09-924-256A-79

  Query Match             30.1%;  Score 30.4;  DB 10;  Length 761;
  Best Local Similarity   57.3%;  Pred. No. 3.5;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       6 ACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGC 65
           ||   || || |||| | | || |    ||  |||         ||  | || || ||||
Db     490 ACTTCCTGTCTCCCGACAGCCGTTCTTGGAGTTTCGATCAACGTGCTAATGGTTATGCGC 549

Qy      66 GGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           | |  || ||   ||||||| ||||  |||| || |
Db     550 GTGGGGAAGGAACCGCCGTGATGGTCGTCAAACGCC 585


RESULT 15
US-09-954-456-694
; Sequence 694, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 694
;   LENGTH: 1359
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-694

  Query Match             30.1%;  Score 30.4;  DB 10;  Length 1359;
  Best Local Similarity   61.2%;  Pred. No. 3.4;
  Matches   49;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           |||||| |   | |  |||||||||| || ||| | ||  |  |  ||  || | || |||
Db     261 GATGCAGGGCTTGTACCCCGGCGGGGGGGGCATGGCGGGCCAGAGCGCGGCCGGCGTCTA 320

Qy      61 CGCGCGGGCCGAGGGCTGCG 80
           ||||   | | | ||   ||
Db     321 CGCGGCCGGCTATGGGCTCG 340


Search completed: November  5, 2002, 23:00:06
Job time : 97.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450
Perfect score: 101
Sequence: 1 gatgcacgcgctttcgcccg.....cgtggtggtgctcaagcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
2	99.4	98.4	68750	3	US-09-335-409-1	Sequence 1, Appli
3	99.4	98.4	68750	4	US-09-568-102-1	Sequence 1, Appli
4	99.4	98.4	68750	4	US-09-567-969-1	Sequence 1, Appli
5	99.4	98.4	68750	4	US-09-568-480-1	Sequence 1, Appli
6	99.4	98.4	68750	4	US-09-568-486-1	Sequence 1, Appli
7	99.4	98.4	68750	4	US-09-568-472-1	Sequence 1, Appli
8	99.4	98.4	68750	4	US-09-567-899-1	Sequence 1, Appli
c 9	62.6	62.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 10	62.6	62.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
11	61.6	61.0	33529	4	US-09-144-085-3	Sequence 3, Appli
c 12	59.4	58.8	333	4	US-09-060-756-318	Sequence 318, App
c 13	56.4	55.8	507	4	US-09-060-756-680	Sequence 680, App
14	56.4	55.8	15872	4	US-09-105-537-1	Sequence 1, Appli
15	56.4	55.8	43280	2	US-08-804-227C-1	Sequence 1, Appli
16	54.6	54.1	11219	1	US-07-642-734C-1	Sequence 1, Appli
17	54.6	54.1	11219	3	US-08-439-009A-1	Sequence 1, Appli
18	54.2	53.7	28958	1	US-08-258-261B-6	Sequence 6, Appli
19	54.2	53.7	28958	1	US-08-456-837-6	Sequence 6, Appli
20	54.2	53.7	28958	1	US-08-457-342-6	Sequence 6, Appli
21	54.2	53.7	28958	1	US-08-457-646A-6	Sequence 6, Appli
22	54.2	53.7	28958	1	US-08-458-076A-6	Sequence 6, Appli
23	54.2	53.7	28958	1	US-08-764-233A-4	Sequence 4, Appli
24	54.2	53.7	28958	1	US-08-457-335A-6	Sequence 6, Appli
25	54.2	53.7	28958	1	US-08-729-214-6	Sequence 6, Appli
26	54.2	53.7	28958	3	US-09-028-934-6	Sequence 6, Appli
27	54.2	53.7	49377	1	US-08-764-233A-1	Sequence 1, Appli
28	53.2	52.7	20235	1	US-07-642-734C-3	Sequence 3, Appli
29	53.2	52.7	20235	3	US-08-439-009A-3	Sequence 3, Appli
30	52.4	51.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
31	52.4	51.9	44377	2	US-08-804-198-1	Sequence 1, Appli
32	52.2	51.7	13987	2	US-08-804-227C-13	Sequence 13, Appl
33	50.8	50.3	751	3	US-09-010-809-4	Sequence 4, Appli
34	50	49.5	80161	3	US-09-036-987A-1	Sequence 1, Appli
35	50	49.5	80161	4	US-09-370-700-1	Sequence 1, Appli
36	47.8	47.3	50937	4	US-09-428-517-1	Sequence 1, Appli
37	46.8	46.3	423	4	US-09-144-085-5	Sequence 5, Appli
38	46.8	46.3	13842	4	US-09-105-537-30	Sequence 30, Appl
39	46.8	46.3	36778	4	US-09-105-537-5	Sequence 5, Appli
40	46.8	46.3	38506	3	US-09-320-878-19	Sequence 19, Appl
41	43.6	43.2	4041	4	US-09-105-537-36	Sequence 36, Appl
42	43.6	43.2	11220	4	US-09-105-537-32	Sequence 32, Appl
43	42	41.6	1462	4	US-09-434-288-4	Sequence 4, Appli
44	42	41.6	4689	4	US-09-105-537-34	Sequence 34, Appl
45	41.8	41.4	658	4	US-08-861-774E-69	Sequence 69, Appl

ALIGNMENTS

```
RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21350 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 21409

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
            |||||||||||||||||||||||||||||||||||||||||
Db   21410 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 21450


RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409
```

```
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             98.4%;  Score 99.4;  DB 3;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 4
US-09-567-969-1
```

```
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 3.6e-16;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             62.0%;  Score 62.6;  DB 4;  Length 4403765;
  Best Local Similarity   79.6%;  Pred. No. 1.5e-07;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
```

```
          | |||  |||||| ||| |||  ||||  |||||| |||||||| ||||||||| | ||
Db 4252406 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4252347

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           |||| ||| ||| | || |||||||||| ||||
Db 4252346 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 4252314


RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match            62.0%;  Score 62.6;  DB 4;  Length 4411529;
  Best Local Similarity   79.6%;  Pred. No. 1.5e-07;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           | |||  |||||| ||| |||  |||| | ||||| |||||||| ||||||||| | ||
Db 4260150 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4260091

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           |||| ||| ||| | || ||||||||||||||
Db 4260090 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 4260058


RESULT 11
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match            61.0%;  Score 61.6;  DB 4;  Length 33529;
  Best Local Similarity   76.0%;  Pred. No. 3.6e-07;
  Matches   76;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 ATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTAC 61
           | ||  ||| | ||||||| ||| ||||||||||||||||||    |  || |||||||||
Db  10349 ACGCGTGCGGTCTCGCCCGACGGACGGTGCAAGACGTTCTCGCAGGCGGCCGACGGCTAC 10408

Qy      62 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |  || | |||||||  ||  |||||||||| | ||| |
Db  10409 GGCCGCGGCGAGGGCGTCGGGGTGGTGGTGCTGATGCGTC 10448


RESULT 12
US-09-060-756-318/c
; Sequence 318, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 318
;   LENGTH: 333
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-060-756-318

  Query Match            58.8%;  Score 59.4;  DB 4;  Length 333;
  Best Local Similarity   77.4%;  Pred. No. 1.6e-06;
  Matches   72;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     182 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 123

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           | |||||||||| ||||||||||||||||||||
Db     122 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGG 90


RESULT 13
US-09-060-756-680/c
; Sequence 680, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/09/060,756
;  CURRENT FILING DATE: 1998-04-16
;  NUMBER OF SEQ ID NOS: 743
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 680
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (various positions within the sequence)
;   OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-680

  Query Match            55.8%;  Score 56.4;  DB 4;  Length 507;
```

```
  Best Local Similarity   74.2%;  Pred. No. 8.2e-06;
  Matches   69;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||  || ||| | |  ||    ||| ||||||| || | || |||
Db     204 GCGCTGTCACCGACCGACCGATGCNACAGNTTNGACGCCAACGCGGATGGATTCGTGCGC 145

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           | |||||||||||| ||||||||||||||||||
Db     144 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGG 112


RESULT 14
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             55.8%;  Score 56.4;  DB 4;  Length 15872;
  Best Local Similarity   73.5%;  Pred. No. 6.5e-06;
  Matches   72;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           | | | |||  |  ||| ||| || |||||| |||||||||||| ||| ||||||  ||
Db    9142 GAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGG 9201

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             ||||||||||||  || |||| | || ||  ||||||
Db    9202 CTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGC 9239


RESULT 15
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             55.8%;  Score 56.4;  DB 2;  Length 43280;
  Best Local Similarity   73.5%;  Pred. No. 6e-06;
  Matches   72;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       4 GCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGC 63
           | | | |||  |  ||| ||| || |||||| |||||||||||| ||| ||||||  ||
Db   37004 GAAGGGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGG 37063

Qy      64 GCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
             ||||||||||||  || |||| | || ||  ||||||
Db   37064 CTGGGCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGC 37101


Search completed: November  5, 2002, 14:35:02
Job time : 1676.74 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450
Perfect score: 101
Sequence: 1 gatgcacgcgctttcgcccg.....cgtggtggtgctcaagcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	71989	21	AAA29349	Sorangium cellulos
2		99.4	98.4	68750	21	AAZ55887	Sorangium cellulos
c 3		62.6	62.0	4403765	22	AAI99683	Mycobacterium tube
c 4		62.6	62.0	4411529	22	AAI99682	Mycobacterium tube
5		61.6	61.0	33529	23	AAS17367	DNA sequence of S.
c 6		59.4	58.8	333	21	ABQ62819	Mycobacterium tube
7		59.4	58.8	5481	22	AAF31638	Mycobacterium tube
8		59.4	58.8	5485	22	AAH52061	Mycobacterium tube
9		57	56.4	758	22	AAF89970	Partial nucleotide
10		57	56.4	5088	22	AAF90038	Nucleotide sequenc
11		57	56.4	34071	22	AAF90033	Nucleotide sequenc
c 12		57	56.4	42717	22	AAF90032	Nucleotide sequenc
13		56.8	56.2	764	22	AAF89974	Partial nucleotide
14		56.8	56.2	6564	22	AAF31637	Mycobacterium tube
c 15		56.4	55.8	507	21	ABQ63181	Mycobacterium tube
16		56.4	55.8	15872	18	AAT68715	Streptomyces venez
17		56.4	55.8	15872	21	AAZ87283	S. venezuelae vep
18		56.4	55.8	43280	18	AAT80413	Tylactone synthase
19		55.2	54.7	671	22	AAF89977	Partial nucleotide
20		54.8	54.3	53789	19	AAV21187	Amycolatopsis medi
21		54.6	54.1	29879	14	AAQ46806	eryA region of S.
c 22		54.2	53.7	14071	23	ABL50560	Micromonospora car
23		54.2	53.7	28598	17	AAT06769	Sorangium cellulos
24		54.2	53.7	28958	18	AAT89956	Sorangium cellulos
25		54.2	53.7	28958	21	AAA75299	DNA sequence of So
26		54.2	53.7	49377	19	AAV05287	The soraphen biosy
27		54	53.5	109519	22	AAS08693	Micromonospora DNA
28		53	52.5	5292	22	AAF90036	Nucleotide sequenc
29		52.4	51.9	44377	18	AAT78508	Platenolide syntha
30		52.4	51.9	44377	18	AAT80414	Platenolide syntha
31		52.4	51.9	47981	22	AAF30757	Micromonospora meg
32		52.2	51.7	13987	18	AAT80415	Hybrid srmG/tylG O
33		52	51.5	668	22	AAF89976	Partial nucleotide
34		52	51.5	65140	22	AAD17184	Streptomyces nours
35		52	51.5	125401	22	AAD17186	Streptomyces nours
36		51.6	51.1	27541	22	AAD17185	Streptomyces nours
37		51.6	51.1	77536	21	AAA14651	Nucleotide sequenc
38		51	50.5	671	22	AAF89969	Partial nucleotide
39		51	50.5	4306	22	AAF90039	Nucleotide sequenc
40		50.8	50.3	671	22	AAF89966	Partial nucleotide
41		50.8	50.3	751	21	AAA58914	DNA encoding ketos
42		50.4	49.9	6333	22	AAF31641	Mycobacterium tube
43		50	49.5	14775	22	AAF88338	S. spinosa DNA fra
44		50	49.5	20394	22	AAF24892	Pimaricin biosynth
45		50	49.5	50000	22	AAF88313	S. spinosa DNA fra

ALIGNMENTS

RESULT 1
AAA29349
ID AAA29349 standard; DNA; 71989 BP.
XX
AC AAA29349;
XX
DT 12-SEP-2000 (first entry)
XX
DE Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /*tag= c
FT /label= epoA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548

```
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                         domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                         domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                         domain, potentially involved in formation of the
FT                         thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                         domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                         module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                         NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                         module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                         module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                         module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                         module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 9e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21350 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 21409

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  21410 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 21450


RESULT 2
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
```

```
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR          AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR          AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR          AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             98.4%;  Score 99.4;  DB 21;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.2e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 3
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             62.0%;  Score 62.6;  DB 22;  Length 4403765;
  Best Local Similarity   79.6%;  Pred. No. 8.9e-07;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           | |||  |||||| ||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4252406 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4252347

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
             |||| ||| ||| | || ||||||||||||||
Db 4252346 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 4252314


RESULT 4
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
```

```
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match             62.0%;  Score 62.6;  DB 22;  Length 4411529;
  Best Local Similarity   79.6%;  Pred. No. 8.9e-07;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCCGCGG 67
           | |||  |||||| ||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db 4260150 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 4260091

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           |||| ||| ||| | || ||||||||||||||
Db 4260090 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 4260058


RESULT 5
AAS17367
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   12-MAR-2002  (first entry)
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
XX
OS   Sorangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT   -
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the DNA
CC   sequence of cosmid pKOS28-26 which encodes one or more domains of
CC   S. cellulosum PKS.
CC   Note: The present sequence is said to encode the functional domains
CC   of S. cellulosum PKS which correspond to domains or domain subsets of
CC   the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

  Query Match             61.0%;  Score 61.6;  DB 23;  Length 33529;
  Best Local Similarity   76.0%;  Pred. No. 2.2e-06;
  Matches   76;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       2 ATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTAC 61
           | ||  ||| | ||||||| ||| ||||||||||||||||||    |  || |||||||||
Db  10349 ACGCGTGCGGTCTCGCCCGACGGACGGTGCAAGACGTTCTCGCAGGCGGCCGACGGCTAC 10408

Qy      62 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |  || | |||||||  ||  ||||||||||| | ||| |
Db  10409 GGCCGCGGCGAGGGCGTCGGGGTGGTGGTGCTGATGCGTC 10448


RESULT 6
ABQ62819/c
ID   ABQ62819 standard; DNA; 333 BP.
XX
AC   ABQ62819;
XX
DT   16-AUG-2002  (first entry)
XX
DE   Mycobacterium tuberculosis BAC vector clone Rv272T7.
XX
KW   Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
KW   detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW   gene; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9954487-A2.
XX
PD   28-OCT-1999.
XX
PF   16-APR-1999;   99WO-IB00740.
XX
PR   16-APR-1998;   98US-0060756.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Cole S,  Buchrieser-Brosch R,  Gordon S,  Billault A;
XX
DR   WPI; 2000-013262/01.
XX
PT   Isolation of polynucleotides from mycobacterial genomes, useful for
PT   detection of Mycobacteria and for combating tuberculosis -
XX
PS   Claim 23; Page 77; 161pp; English.
```

```
XX
CC   The present invention describes a method for isolating a polynucleotide
CC   of interest that is present or is expressed in a genome of a first
CC   mycobacterium strain and that is absent or altered in a genome of a
CC   second mycobacterium strain, which is different from the first strain
CC   using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
CC   vectors, which are preferably immobilised, can be used to detect
CC   mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
CC   samples. The polynucleotides identified are useful as probes or primers
CC   for detecting a given mycobacterium of interest. By aligning the
CC   polynucleotides contained in the recombinant BAC vectors it is possible
CC   to physically map a polynucleotide of mycobacterial origin in a
CC   biological sample. The methods and vectors from the present invention
CC   are useful in providing information for combating tuberculosis. It is
CC   possible to compare genomes between different strains or species and
CC   their non-pathogenic strains or species counterparts. ABQ62492 to
CC   ABQ63228 and ABB81227 to ABB81230 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 333 BP; 66 A; 117 C; 102 G; 48 T; 0 other;

  Query Match             58.8%;  Score 59.4;  DB 21;  Length 333;
  Best Local Similarity   77.4%;  Pred. No. 1e-05;
  Matches   72;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGAC3TTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     182 GCGCTGTCACCGACCGGCCGATGCAACAG:TTCGACGCCAACGCGGATGGATTCGTGCGC 123

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAA3CGG 100
           | |||||||||||| ||||||||||||||||||
Db     122 GGCGAGGGCTGCGGCGTGGTGGTGCTCAA:CGG 90


RESULT 7
AAF31638
ID   AAF31638 standard; DNA; 5481 BP.
XX
AC   AAF31638;
XX
DT   09-APR-2001  (first entry)
XX
DE   Mycobacterium tuberculosis ppsD gene.
XX
KW   Mycobacterium tuberculosis; attenuated microorganism;
KW   signature tagged transposon mutant; mutant library;
KW   mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW   vaccine; ppsD; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200102555-A1.
XX
PD   11-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-IB00950.
XX
PR   06-JUL-1999;   99US-0142982.
PR   08-JUL-1999;   99US-0142833.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Guilhot C,  Camacho L;
XX
DR   WPI; 2001-091804/10.
DR   P-PSDB; AAB66468.
XX
PT   Screening a mutant library for mutants unable to grow under specific
PT   conditions and for identifying loci involved in pathogenicity,
PT   comprises using signature tagged transposon mutagenesis  -
XX
PS   Example 8; Page 131-132; 159pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   screening a library of mutants. The method comprises constructing a
CC   library with insertions in genes and/or regulatory regions of the
CC   organisms of interest, where the insertion contains a tag and/or a
CC   transposon associated with a tag. The mutants are identified by
CC   hybridisation of the tags to known sequences. The method is useful for
CC   treating an individual suffering from a mycobacterial infection,
CC   suspected of being infected with a Mycobacterium, or having been
CC   exposed to an infectious Mycobacterium. It is also useful for
CC   identifying and isolating mutants of actinomycetales and for identifying
CC   compounds that have antibiotic activity. The method is used to identify
CC   mutants of microorganisms, preferably an actinomycetales, such as
CC   M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC   M. paratuberculosis, that is unable to grow under specific conditions.
CC   It is especially useful for identifying loci involved in pathogenicity.
CC   It is useful in constructing vaccines. The method can be used to screen
CC   multiple libraries concurrently. It can screen libraries of different
CC   organisms or different strains of the same organism.
XX
SQ   Sequence 5481 BP; 894 A; 1745 C; 1925 G; 917 T; 0 other;

  Query Match             58.8%;  Score 59.4;  DB 22;  Length 5481;
  Best Local Similarity   77.4%;  Pred. No. 8.4e-06;
  Matches   72;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     751 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 810

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           | |||||||||||| ||||||||||||||||||
Db     811 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGG 843


RESULT 8
AAH52061
ID   AAH52061 standard; DNA; 5485 BP.
XX
AC   AAH52061;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Mycobacterium tuberculosis potential drug target gene SEQ ID 115.
XX
KW   Drug target; growth; organism viability; characterisation; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200135317-A1.
XX
PD   17-MAY-2001.
XX
PF   13-NOV-2000; 2000WO-US31152.
XX
PR   12-NOV-1999;   99US-0165086.
PR   12-NOV-1999;   99US-0165124.
PR   01-FEB-2000; 2000US-0179531.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Eisenberg D,  Rotstein SH,  Marcotte EM;
XX
DR   WPI; 2001-329193/34.
DR   P-PSDB; AAG81210.
XX
PT   Identifying nucleotide or polypeptide sequence for use as drug target,
PT   involves providing algorithm that analyzes a functional relationship
PT   between nucleotide or polypeptide sequences, and comparing the
PT   sequences  -
XX
PS   Disclosure; Page 128-130; 207pp; English.
XX
CC   This invention relates to a method for identifying a nucleotide or
```

```
CC   polypeptide sequence that may be a drug target, or essential for growth
CC   or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC   represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC   tuberculosis proteins which are potential drug targets. The DNA and
CC   protein sequences are used to illustrate the method of the invention. The
CC   method involves providing an unknown nucleotide or polypeptide sequences,
CC   and comparing it to a number of sequences along with at least one
CC   algorithm capable of analysing a functional relationship between
CC   nucleotide and polypeptide sequences. The method is useful for
CC   characterising the function of nucleic acids and polypeptides that may be
CC   useful as a target for a drug or essential for the growth or viability of
CC   an organism.
XX
SQ   Sequence 5485 BP; 896 A; 1745 C; 1926 G; 918 T; 0 other;

  Query Match             58.8%;  Score 59.4;  DB 22;  Length 5485;
  Best Local Similarity   77.4%;  Pred. No. 8.4e-06;
  Matches   72;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||| || ||||| |  |||   ||| ||||||| || | || |||
Db     751 GCGCTGTCACCGACCGGCCGATGCAACAGCTTCGACGCCAACGCGGATGGATTCGTGCGC 810

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           | |||||||||||| ||||||||||||||||||
Db     811 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGG 843


RESULT 9
AAF89970
ID   AAF89970 standard; DNA; 758 BP.
XX
AC   AAF89970;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83962.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 222; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 758 BP; 121 A; 250 C; 247 G; 140 T; 0 other;

  Query Match             56.4%;  Score 57;  DB 22;  Length 758;
  Best Local Similarity   74.2%;  Pred. No. 3.6e-05;
  Matches   72;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       5 CACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 64
           | |  | |  |||||| ||| || ||||||||||||   ||||  || ||||| || | |
Db     469 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 528

Qy      65 CGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           || | ||||||||||| | | ||| |||| |||||||
Db     529 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGC 565


RESULT 10
AAF90038
ID   AAF90038 standard; DNA; 5088 BP.
XX
AC   AAF90038;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5088
FT                   /*tag=  a
FT                   /product= "type I polyketide synthase"
FT                   /transl_except= "(1..3, aa: Met)"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83975.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 309-311; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
```

```
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 5088 BP; 838 A; 1622 C; 1679 G; 949 T; 0 other;

  Query Match           56.4%;  Score 57;  DB 22;  Length 5088;
  Best Local Similarity  74.2%;  Pred. No. 3.1e-05;
  Matches   72;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       5 CACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 64
           | |  | |  |||||| ||| || |||| |||||||   ||||  || ||||| || | |
Db     739 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 798

Qy      65 CGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           || | ||||||||||| | | ||| |||| |||||||
Db     799 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGC 835


RESULT 11
AAF90033
ID   AAF90033 standard; DNA; 34071 BP.
XX
AC   AAF90033;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Example 14; Page 289-300; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 coding strand), which encodes different
CC   polyketide synthases.
XX
SQ   Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

  Query Match           56.4%;  Score 57;  DB 22;  Length 34071;
  Best Local Similarity  74.2%;  Pred. No. 2.7e-05;
  Matches   72;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy       5 CACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 64
           | |  | |  |||||| ||| || |||| |||||||   ||||  || ||||| || | |
Db   25420 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 25479

Qy      65 CGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           || | ||||||||||| | | ||| |||| |||||||
Db   25480 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGC 25516


RESULT 12
AAF90032/c
ID   AAF90032 standard; DNA; 42717 BP.
XX
AC   AAF90032;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Example 14; Page 274-288; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC   encodes different polyketide synthases.
XX
SQ   Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

  Query Match           56.4%;  Score 57;  DB 22;  Length 42717;
  Best Local Similarity  74.2%;  Pred. No. 2.7e-05;
  Matches   72;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;
```

```
Qy      5 CACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCG 64
          | |  | |  |||||| ||| || ||||||||||||   ||||  || ||||| || | |
Db   8700 CGCATGTTGGCGCCCGACGGACGCTGCAAGACGTTCGACGCCGCAGCCGACGGTTATGTG 8641

Qy     65 CGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          || | |||||||||||| | | ||| |||| |||||||
Db   8640 CGCGGCGAGGGCTGCGGCATCGTGCTGCTGAAGCGGC 8604
```

RESULT 13
AAF89974
ID AAF89974 standard; DNA; 764 BP.
XX
AC AAF89974;
XX
DT 06-AUG-2001 (first entry)
XX
DE Partial nucleotide sequence of a type I polyketide synthase.
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
XX
OS Unidentified.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
DR P-PSDB; AAB83966.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics -
XX
PS Claim 35; Page 224; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC encode partial type I polyketide synthases, and were isolated using the
CC method of the invention.
XX
SQ Sequence 764 BP; 114 A; 272 C; 268 G; 110 T; 0 other;

Query Match 56.2%; Score 56.8; DB 22; Length 764;
Best Local Similarity 76.1%; Pred. No. 4e-05;
Matches 70; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy     10 GCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGC 69
          |||  |||| | |||||| ||||||||||||   ||||  ||||||||||||  || |
Db    480 GCTCGCGCCTGACGGGCGCTGCAAGACGTTCGACGCCGCGGCGGACGGCTACGTCCGCGG 539

Qy     70 CGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          ||| ||   ||| ||| | ||||| |||||||
Db    540 CGAAGGGGCCGCGGTGATCGTGCTGAAGCGGC 571
```

RESULT 14
AAF31637
ID AAF31637 standard; DNA; 6564 BP.
XX
AC AAF31637;
XX
DT 09-APR-2001 (first entry)
XX
DE Mycobacterium tuberculosis ppsC gene.
XX
KW Mycobacterium tuberculosis; attenuated microorganism;
KW signature tagged transposon mutant; mutant library;
KW mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW vaccine; ppsC; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200102555-A1.
XX
PD 11-JAN-2001.
XX
PF 06-JUL-2000; 2000WO-IB00950.
XX
PR 06-JUL-1999; 99US-0142982.
PR 08-JUL-1999; 99US-0142833.
XX
PA (INSP) INST PASTEUR.
XX
PI Gicquel B, Guilhot C, Camacho L;
XX
DR WPI; 2001-091804/10.
DR P-PSDB; AAB66467.
XX
PT Screening a mutant library for mutants unable to grow under specific
PT conditions and for identifying loci involved in pathogenicity,
PT comprises using signature tagged transposon mutagenesis -
XX
PS Example 8; Page 122-124; 159pp; English.
XX
CC The present sequence is given in a specification relating to a method for
CC screening a library of mutants. The method comprises constructing a
CC library with insertions in genes and/or regulatory regions of the
CC organisms of interest, where the insertion contains a tag and/or a
CC transposon associated with a tag. The mutants are identified by
CC hybridisation of the tags to known sequences. The method is useful for
CC treating an individual suffering from a mycobacterial infection,
CC suspected of being infected with a Mycobacterium, or having been
CC exposed to an infectious Mycobacterium. It is also useful for
CC identifying and isolating mutants of actinomycetales and for identifying
CC compounds that have antibiotic activity. The method is used to identify
CC mutants of microorganisms, preferably an actinomycetales, such as
CC M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC M. paratuberculosis, that is unable to grow under specific conditions.
CC It is especially useful for identifying loci involved in pathogenicity.
CC It is useful in constructing vaccines. The method can be used to screen
CC multiple libraries concurrently. It can screen libraries of different
CC organisms or different strains of the same organism.
XX
SQ Sequence 6564 BP; 1137 A; 2189 C; 2154 G; 1084 T; 0 other;

Query Match 56.2%; Score 56.8; DB 22; Length 6564;
Best Local Similarity 76.1%; Pred. No. 3.4e-05;
Matches 70; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy     10 GCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGGGC 69
          ||| || || |  |||||||||||||| |||   ||   ||| || || |||| ||| |
Db    762 GCTGTCACCGGAGGGGCGGTGCAAGACCTTCGATGCGTCCGCCGATGGATACGTGCGCGG 821
```

```
Qy      70 CGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           ||||||   ||| ||||||||||||||||||
Db     822 CGAGGGTGCCGCGGTGGTGGTGCTCAAGCGGC 853


RESULT 15
ABQ63181/c
ID   ABQ63181 standard; DNA; 507 BP.
XX
AC   ABQ63181;
XX
DT   16-AUG-2002  (first entry)
XX
DE   Mycobacterium tuberculosis BAC vector clone Rv90T7.
XX
KW   Mycobacterium tuberculosis; Mycobacterium bovis; mycobacterium;
KW   detection; BAC vector; bacterial artificial chromosome; tuberculosis;
KW   gene; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9954487-A2.
XX
PD   28-OCT-1999.
XX
PF   16-APR-1999;   99WO-IB00740.
XX
PR   16-APR-1998;   98US-0060756.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Cole S,  Buchrieser-Brosch R,  Gordon S,  Billault A;
XX
DR   WPI; 2000-013262/01.
XX
PT   Isolation of polynucleotides from mycobacterial genomes, useful for
PT   detection of Mycobacteria and for combating tuberculosis -
XX
PS   Claim 23; Page 121; 161pp; English.
XX
CC   The present invention describes a method for isolating a polynucleotide
CC   of interest that is present or is expressed in a genome of a first
CC   mycobacterium strain and that is absent or altered in a genome of a
CC   second mycobacterium strain, which is different from the first strain
CC   using a bacterial artificial chromosome (BAC) vector. Recombinant BAC
CC   vectors, which are preferably immobilised, can be used to detect
CC   mycobacterial nucleic acids (genomic DNA, cDNA or mRNA) in biological
CC   samples. The polynucleotides identified are useful as probes or primers
CC   for detecting a given mycobacterium of interest. By aligning the
CC   polynucleotides contained in the recombinant BAC vectors it is possible
CC   to physically map a polynucleotide of mycobacterial origin in a
CC   biological sample. The methods and vectors from the present invention
CC   are useful in providing information for combating tuberculosis. It is
CC   possible to compare genomes between different strains or species and
CC   their non-pathogenic strains or species counterparts. ABQ62492 to
CC   ABQ63228 and ABB81227 to ABB81230 represent sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence 507 BP; 89 A; 176 C; 132 G; 87 T; 23 other;

  Query Match            55.8%;  Score 56.4;  DB 21;  Length 507;
  Best Local Similarity  74.2%;  Pred. No. 5.2e-05;
  Matches   69;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           ||||| || ||   ||  || ||| | |  ||    ||| ||||||| || | || |||
Db     204 GCGCTGTCACCGACCGACCGATGCNACAGNTTNGACGCCAACGCGGATGGATTCGTGCGC 145

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
           | ||||||||||| ||||||||||||||||||
Db     144 GGCGAGGGCTGCGGCGTGGTGGTGCTCAAGCGG 112
```

```
Search completed: November  5, 2002, 13:04:56
Job time : 1762.22 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_21350_21450
Perfect score: 101
Sequence: 1 gatgcacgcgctttcgcccg.....cgtggtggtgctcaagcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	71989	6	AR172664	AR172664 Sequence
3		99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4		99.4	98.4	68750	6	AR193029	AR193029 Sequence
5		99.4	98.4	68750	6	AR199551	AR199551 Sequence
6		99.4	98.4	68750	6	AR199559	AR199559 Sequence
7		99.4	98.4	68750	6	AR199567	AR199567 Sequence
8		99.4	98.4	68750	6	AR201097	AR201097 Sequence
9		99.4	98.4	68750	6	AR208671	AR208671 Sequence
10		80.8	80.0	10910	1	AX024383	AX024383 Sequence
11		80.8	80.0	10910	6	AX024276	AX024276 Sequence
12	c	64.2	63.6	27522	1	AB070942	AB070942 Streptomy
13		63.8	63.2	66808	1	SAU421825	AJ421825 Stigmatel
14	c	62.6	62.0	14634	1	AE007184	AE007184 Mycobacte
15	c	62.6	62.0	23740	1	MTV026	AL022076 Mycobacte
16		62.2	61.6	42603	1	AF188287	AF188287 Stigmatel
17		61.6	61.0	33529	6	AR166425	AR166425 Sequence
18	c	61.2	60.6	82746	1	AF453501	AF453501 Actinosyn
19	c	59.6	59.0	30000	6	AX250263	AX250263 Sequence
20	c	59.4	58.8	333	6	AR128456	AR128456 Sequence
21		59.4	58.8	5481	6	AX069075	AX069075 Sequence
22		59.4	58.8	20679	1	MTCY19H9	Z83857 Mycobacteri
23		59.4	58.8	32870	1	AE007122	AE007122 Mycobacte
24	c	59.2	58.6	19601	1	AB070944	AB070944 Streptomy
25	c	58.8	58.2	49736	1	AF319998	AF319998 Stigmatel
26		58.4	57.8	213050	1	AL646079	AL646079 Ralstonia
27		58.2	57.6	90348	1	AF497482	AF497482 Micromono
28		57	56.4	758	6	AX153712	AX153712 Sequence
29		57	56.4	5088	6	AX153795	AX153795 Sequence
30		57	56.4	34071	6	AX153790	AX153790 Sequence
31	c	57	56.4	42717	6	AX153789	AX153789 Sequence
32		56.8	56.2	764	6	AX153716	AX153716 Sequence
33		56.8	56.2	6564	6	AX069073	AX069073 Sequence
34	c	56.4	55.8	507	6	AR128818	AR128818 Sequence
35		56.4	55.8	3185	1	AF411573S4	AF411576 Actinomad
36		56.4	55.8	3900	1	AF262754	AF262754 Amycolato
37		56.4	55.8	43280	1	SFU78289	U78289 Streptomyce
38		56.4	55.8	78210	1	AB070949	AB070949 Streptomy
39		55.6	55.0	3573	1	AB016763	AB016763 Streptomy
40		55.2	54.7	671	6	AX153719	AX153719 Sequence
41	c	55.2	54.7	189050	1	AL646066	AL646066 Ralstonia
42		54.8	54.3	53784	1	AMM223012	AJ223012 Amycolato
43		54.8	54.3	53789	6	A69720	A69720 Sequence 3
44		54.8	54.3	90445	1	AF040570	AF040570 Amycolato
45		54.6	54.1	11219	1	SERERYAA	M63676 Saccharopol

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"

CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"

CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLRPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"

gene 1998. .6263
/gene="epoA"

CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRFIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPEGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"

gene 6260. .10492
/gene="epoB"

CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"

gene 10639. .16137
/gene="epoC"

CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARGIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 3e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21350 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 21409

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   21410 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 21450


RESULT 2
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match          100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21350 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 21409

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   21410 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 21450


RESULT 3
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
```

/db_xref="GI:6724239"
/translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
SSPIQRAIDELGLVNACPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
IGIVMTQSAGFLFSGALERFWRSVYVATESA"
CDS 3415. .5556
/note="ORF3; similar to Na/H and drug/H antiporters"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF25906.1"
/db_xref="GI:6724240"
/translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLRVGIWRQRRELRGS
IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
ARSVVDEAVFRGRRLGVEVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
YLGSTVESVVVRSRVPVLLVAHGGTREQVR"
CDS complement(5612. .5992)
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26907.1"
/db_xref="GI:6724241"
/translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
QHTRLDSRQAENALLSTIAYTDRE"
CDS 6226. .6675
/note="ORF5"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF26908.1"
/db_xref="GI:6724242"
/translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
YNADSRRGKLAETTVPYAWFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene 7610. .11875
/gene="epoA"
CDS 7610. .11875
/gene="epoA"
/note="EPOS A"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26919.1"
/db_xref="GI:6724253"
/translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG

VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature 7640. .8920
/gene="epoA"
/note="Region: beta-ketoacyl synthase/malonyl-CoA
decarboxylase"
misc_feature 9236. .10201
/gene="epoA"
/note="Region: acetate acyltransferase"
misc_feature 10529. .11428
/gene="epoA"
/note="Region: enoyl reductase"
misc_feature 11549. .11764
/gene="epoA"
/note="Region: acyl carrier protein"
gene 11872. .16104
/gene="epoP"
CDS 11872. .16104
/gene="epoP"
/note="EPOS P; contains condensation/heterocyclization,
cysteine adenylation, putative NAD(P)H oxidase, and
peptidyl carrier domains"
/codon_start=1
/transl_table=11
/product="nonribosomal peptide synthetase"
/protein_id="AAF26925.1"
/db_xref="GI:6724259"
/translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene 16251. .21749
/gene="epoB"
CDS 16251. .21749
/gene="epoB"
/note="EPOS B"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="AAF26920.1"
/db_xref="GI:6724254"

Query Match 98.4%; Score 99.4; DB 1; Length 68750;

```
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 4
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US-6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 5
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 6
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 7
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062
```

```
RESULT 8
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison , Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 9
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5.6e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26962 GATGCACGCACTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 27021

Qy     61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db  27022 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 27062


RESULT 10
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           80.0%;  Score 80.8;  DB 1;  Length 10910;
  Best Local Similarity   88.0%;  Pred. No. 2.5e-06;
  Matches   88;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 ATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTAC 61
          ||||    ||||||||| |  ||||||||||||||||||||||  || |||||||
Db   2895 ATGCGTTTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGGCTTT 2954

Qy     62 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   2955 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 2994


RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           80.0%;  Score 80.8;  DB 6;  Length 10910;
  Best Local Similarity   88.0%;  Pred. No. 2.5e-06;
  Matches   88;  Conservative    0;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      2 ATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTAC 61
          ||||    ||||||||| |  ||||||||||||||||||||||  || |||||||
Db   2895 ATGCGTTTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGGCTTT 2954

Qy     62 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db   2955 GCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 2994


RESULT 12
AB070942/c
LOCUS       AB070942               27522 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyketide-2 biosynthetic gene cluster.
ACCESSION   AB070942
VERSION     AB070942.1  GI:15823997
```

```
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 27522)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .27522
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .555)
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB69213.1"
                     /db_xref="GI:15823998"
                     /translation="MAEETAAPKADAAAPRSFTLSDTTAGPVGTRGTRVFIDKQSPKA
                     YHALVETSEAVRAVAADAGLDRVLVELINLRVSQINGCAYCLSLHTKAALRVGETTQR
                     LGVLAAWRDTELFTSRERAALALAEVTTDPADATAQEHAYDAARQVLTDDEISAVIWV
                     AVTINAFNRVSIMSKHPVRGIPHQ"
     CDS             complement(584. .1030)
                     /codon_start=1
                     /transl_table=11
                     /product="putative isomerase"
                     /protein_id="BAB69214.1"
                     /db_xref="GI:15823999"
                     /translation="MDEPVARSLLAELRGRGHEVVAHGALRPGDDTRWAACSAAAARE
                     VATGTSDQAVVCCWTGTGASIAANKVPGVRAALCTDAYTADGARRWNDANVLALSLRL
                     TSQPLLREILDAWFAAEAGEDADDRENVAHVGRLDADRTGGAADRG"
     CDS             complement(1069. .1869)
                     /codon_start=1
                     /transl_table=11
                     /product="putative carboxylase"
                     /protein_id="BAB69215.1"
                     /db_xref="GI:15824000"
                     /translation="MPQLAVDGAGMTYDDEGPHDGDGGVPLVFIHGWTADRHRWDHQM
                     AHFADKRRVVRLDLRGHGESGGSARTIDELAGDVIALLDHLEIDRFIPVGHSMGGMIA
                     QTLALAHPERIERLVLVNSISRMTYSRGRGLLMAASTLVPFKLFVAANIQRAFAPGHP
                     REEVREYIRASSATPRDVVMTYYAAMRSFDVLDRVGEIRMPTLLVHGYYDIQLPVSQM
                     LRMAKDYPDAVVRIVDAGHELPVEKPAELTSALDRFVTGQVGGVQLVG"
     CDS             complement(1933. .3069)
                     /codon_start=1
                     /transl_table=11
                     /product="probable esterase/lipase"
                     /protein_id="BAB69216.1"
                     /db_xref="GI:15824001"
                     /translation="MIFRTTDSTTRPALRRVRTTTTLRPLHHRLDPARVEEIPFRAGD
                     GVRLGLTRVDSGEPDRPAVLLLHGHTASADMFLLPETRNLVDALLDDGYEPWLLDWRG
                     SCRLPYNETGQKYTYDDVALYDIPAAVSHIRGRIGDRPLFVVAHCIGSLTLSLSMTAG
                     LVPGLAGVVSQGVFLTPKLAGRTSLRMSVAGELLKTRIDHIPVDFRKVGLWSKYTPLF
                     ALASRKASCPDPTCQILHNSAWGTGASLFVHEHLSEATHDRLAELLGPAPLWILPHLR
                     RIELARSVVRWHDTDHRYRELPQNALDAAGRIDTPVLLIAGSENGLWLDSQKLCHDVL
                     ARRQPYLDVSYTEIPGYGHLDTFLGRGAALDVFGHILDFLGERR"
     CDS             complement(5628. .20057)
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69217.1"
                     /db_xref="GI:15824002"
                     /translation="MRAWLSAAVADAAGLQAPVIDLDRPIAEFGLSSRQLVTLAAELS
                     QRIDRPLNPSLVFDHPTIAGIAEAVFDESPVGATAADPAPADAPPARADDDIAIVSMA
                     CRFPGGADDPEALWRLLTAGEDAISEVPAGRWDTQGLYDPDPEATGKAYTLRGGFLSD
                     IDRFDASFFGISPREAAAMDPQQRLLLGTAWETIERAGIVPETLNGSSTGVYIGLYDS
                     GYQASASLGRLDGHVGTGSASSVASGRISYTLGLQGPAVTVDTACSSSLVALHLAARA
                     LADGECDLALTGGATLLVTPRGHVEFSRLRGLSPSGRCSPFSTEADGVVWAEGCGLVL
                     LKRLADARRDGDRVLAVVKGSAINQDGRSQGLSAPSGPAQERVLRAALDAAGLRPDDL
                     DYVEAHGTGTPLGDPIEGRALTAVFGPGRPADRPLGVGSLKSNLGHTQAAAGIGGVIK
                     TVLALGHQLLPASLHAETPTEHIDWAGGGLRVHSTAAGWPRDVDRVRRAGVSAFGISG
                     TNAHVVLEEAPHDLTRPTGEPQSPEQPTAEETAPHLPAPHLPAPHLTGKTLFPLSART
                     LPALRGQAGRLLEALSADPQTPLPEAAATLAHHRTHFEHRAVVQAGDRDELLTALRGL
                     AEGQPDTDRVAGPQQALPPGKLAFVFPGQGAQWTGMARDLLDRSPVFADELDRCDAAL
                     RPFTTWSVAAVLRGDAGAPSPDRVDVVQPVLFAVMMSLAAVWRARGVRPDAVIGHSQG
                     EVAAACVAGALSLNDAAAVVALRSQALTELSGTGAMAVVALPHTEVEAHLAGLDGRVS
                     VAAVNSGRSTVIAGDVEPLEALLADLDRQQVFVRRLDVDYASHSAQVEQVRAAVLDEL
                     DGITTCPTSVAWYSTVLGEPVTEELEADYWYTNLREPVRFAPTVERMAADGYRYFVEL
                     SPHPSLLTALRTVVEDADAAEAAEDTAPDLVAVGSLRRDEDGPACLDRAAAELHVHGR
                     RIDWRRLVADAPRADLPTYAWDPQRHWIEPEAAAGAPGLFDRAAHPLLGIQLQSADET
                     RWTFRNEWSPATADWLPDHAVFGRTVVSGTTLMELCRAALAVARPDAPGDVTDLFLLT
                     PLTLPGSGTVEVSVEVVTAGPVPEITVHSRPRGQEGTGWTLHATASAAEAAPVTQGPL
                     PVWPEAAGTAWSPKTYERLARLGLGYGPAFQGVRAAVTTGDGELLARLSLPTVARDVA
                     DPYPLHPAVLDAALHVAAVFDAGDASDGRVLLPVAVGRCVLPHGGATELTARVRRSGG
                     SGTDLTLDVTLWDADGIPAGRLEGVRLRAANPADLGTGSENGRHLYEVAWTAVPEEPA
                     GAPGTEWTVLGDPSDPEAAAALRGLAAAGITVGETGADVRVRFWPRPASDVEPAAAAQ
                     ELAATALAELQALLALPEDEAPTRTVWVTHGAVAAGEGDTVPGLAQSVLWGLARSART
                     EHPGLGLTLLDLDGSDPGGALPVAVAYAGEPELALRAGTLLVPRLVRARAAQALPIPA
                     DDHYELAGLPADVTLRAVAPRALAPGQVRIQVHAARVHPSGDDAPDAVSAGVLTEVGA
                     DVEGLALGLRVVAVADTAPGSAAVADARLVVALPDGLPFTTAVTLPGEHADAVAPGRA
                     ELGAALREAVASVAAGHTRPTRPRAWPVTAAREAIRRGGPVVLDLVGRPTLAPTDGTV
                     LITGGLGAVGRHIARLLAEHGVPRLLLTSRQGADDPRAAEVRAELAALGAEVEVAACD
                     VADAAALADVLGRIGDELPLRGVVHCAGVLADGVVAELTPERLTRVLRPKVDGAAHLH
                     RLTADVPLDLFLLVSSAAGVLGTAGQSNYAAANVFLDQLAHHRRALGLPGVSVSFGAW
                     AGEGLAAEHADLERMARLGHRALTPDQGRELVELSLRRGAPHLVAWSLDLPRLRETTA
                     AGDDTAAALWRSLLPAPRTGQDGADGLADRLARLPEAERAERVLALVREEASRALGLR
                     SAQSVRPDQPLRELGMDSVTAVELRNRIGTRIGAKLPATLLFDHPTPSRLAEYLLTTA
                     LATAGRPARRTAPGAQPARTATPASDEPVAVVSMACRLPGGVSDPEGLWHLVAQGRDA
                     VGPFPAGRWDVESLYDPDPEALGKSYAREGGFLDGIESFDAGFFGITPKEAAAMDPQQ
                     RLLLEMAWEALERAGTVPAELAGSTTGVYVGMFGSDYLAGSRLDQLDGYVGTGSALSV
                     ASGRLSYALGLHGPALTVDTACSSSLVATHLAAQALRSGECDLALAGGVTLMVTPGTF
                     VEFSRLRGLSPTGRCRSFSDDADGAVWAEGAGMLVLKRLSDARRDGDQVLAVLRGTAV
                     NQDGRSQGLSAPNGPAQEQVIRRALELSALGAADIDYVEAHGTGTTLGDPIEANALSE
                     VFGDSRPENRPLYLGSLKSNIGHAQAASGVLGLIKVVQSLRHGTLPRTLHAGTPSQHV
                     DWDGSGLHLLQEAVDWPSSGERVRRAGVSAFGISGTNAHVIVEEAPPPAATEPNAEPM
                     PGKRLFALSGRSEAGVRGQAARLAQYLTEDVALPDVAHTLARHRSHFERRTAIVAADR
                     DELRAALDALATGRTPLVPPREEQTGKVAFVFAGHGGQWPGMGVELMAGSDAFREELA
                     RIDEAVRRHVGWSVLNVLRAPEEIAPLDRTEFLQPVLFAVNAALAAAWRALGITPDAV
                     IGHSLGEIAAAYSAGALTLDAAVTVVTRRAQAVVPLAGKGGMLSVELPRAQVEELLAP
                     YAGRLFVAAVNSAHATAVSGDAEALAEVLHRLEERKIPARRLSTPFASHTPLMDPLRE
                     ELLDRFADVRGARTPTPLYSTVLAEPVPGDRLDAAYWYANLREPVRFADTVRRMLDDG
                     YRYFVELSPHPSLGPSIEAVAAEAGIDAVGVGSLRRQQDGRDILLRRLGELYAAGHTP
                     DWPVLFPTGRRVDLPTYAFARERHWLAPAPASPAGGGGPLLGTHVEASDEPDRHLFQS
                     DLDLRDSRFAYLTDHRVTGDVWLPGAAFLDMALEAASALQDGGEVRLADVRFLRPLRL
                     DEGRPVRLQLVLRPAEDGVRDFTIASAVTGERRTRWERHVSGRVTVTAGEPMADARLA
                     VLREQCTEQVDLPGVYARLAALGIDYGPAFRGLEEGRRTHSAALGRLAAHPAAGHLLH
                     PAVLDAAFHTAALPGDAPEGRAFVPAGFGRLRFTGLRTTPVWVTCELRSMAGDTATLD
                     LRLWDENEQLVLEAEEFELAALSPLDGALFETRWQPRPPAQEPPATGSWLILADETGV
                     AVELAERLGSSVPHVIVRRGEAFGVEGPGRYVLDPADPQQLAHLLDEAFADGLPERVV
                     QLSALDAPAIGEARTAEQAARLCCLSTLHLVRTLADRPEGRAPRLFVVVRGSQAVGDS
                     TQVTHPQQALAWGFGLAVAQEYPELATTLIDLPSTDGVDALWEQLRHADDERLVALRD
                     TGRLVPRLARTRPDDGGHGELTPDGVYLITGGLGGLGRVVAERLVRRGARRLALLSRG
                     RPGADAAGWIKGLEERGVTVHLARADVADRAGLTAALAALRRDAGPVAGVVHAAGVLD
                     DATVANLTDERVLRVLAPKVLGTTLLTELIPEAGNLVLFASAAGLLGSAGQSPYSAAN
                     AFLDAWAHHLSRTDRRALSLDWGAWAGVGMVAESGTRADETSRSGLVAFSPDEGGELF
                     ERVLATSRRQLAPIALDWEMLALDADAARTRPLLGDLVTVPTGGADSDDLVKKVFAAT
                     TDHDRAVRLEAYVRARVGEVCGGAVDVSGTTVLKELGLDSLMLVRLRNAFARELGVEL
                     PAAAMFSASDIRGLVRALVAALPERDTVTPEQPADVTSEVLEVPETELRPATRDVVRL
                     LRSAQPGMPGAAHAVGLAVRLTTPTTRETLTGILTRLVARHAALRTAIVPGAEGGRQL
                     RVDREPANPVLRWTDVAADVELDAGDRLRELLESPFDLTTAALWRFELLDGGARGQIL
                     VFGAHHAVSDLQSLLLVAGEIDAELSGTSLGDTVTNRDIDLLIQAQRTEGTVPTGDGP
                     GAAAAVEWREAFHGSGRLDLTLTRERPAARSYRAGSVTVAIPDGLMERVSAAASGLAV
```

```
                     TPAAFCLGTLTVLLAHRERERFVLAVPVDTRIHADAYDAVGFFGVPVPFPAEAGAGE
                     PVEQVLRRTDGRLEKVLAKGAMFSDVLPTLAKQGLYRPNAPLVEVYFNYVRSSAGRLT
                     NLEVLPAGTGWSDLDLMITMTPDAGRIRLDHNLDILDAAATTDLGEEFLRLLARAVED
                     PTAPVRTTAASEAPAATEVLVASEVPPAAEALAVSEAPAATEVPPATEVPPAGEPAVV
                     PPRPLAVAATFALGNLPLMCELACNEGAGARDGGVTVAEAPYHHVLASLRDPSGVFAD
                     PATAVGVVLLRAADLERFGPVDDALLAELRTEYPAALRALSERTRKPLIVGFPPSARA
                     DDRLERWDSEVAAELAGLPGIAVLGPDDWTRHHAVGERFDERTEKLAHLPFTPDFQAA
                     VALRLADVVRAVRRPAPLVIAVDGDETLWGGVAGEIGPEAVDLTGPRALLARRLLQWR
                     AAGALLALVSNNDEDTVRAVLDRPDSLLKAEHFSVLSAAWGPKSGRLAEAARTLNLGT
                     DSFLFLDDNPVEIAGVRSALPEVLSVTCPPTAELDEFLGRLWPLVPAAATAEDGLRAR
                     FYEQERERDAVREHAGFIEEFLAQLELEVDIRALSAADVQRAEQLVRRTNQFTLRARSA
                     DGGDLTRWRECGEVWTALARDRFGDYGQIGLLAVRPGEGQLDVLAWMMSCRALGRGVE
                     ERLLRWLADRAEELGCALVRLTAERTPRNVPARRLVAALGGGDQDDAYLETVVTPDQL
                     RAFRSWQQQ"
     CDS             complement(20358. .22082)
                     /codon_start=1
                     /transl_table=11
                     /product="non-ribosomal peptide synthetase"
                     /protein_id="BAB69218.1"
                     /db_xref="GI:15824003"
                     /translation="MARSLVDLLTKHASHQPDRTAYRYLVTGDRDGEIQDISYGRLAE
                     RSRAVASWLQERGLAGSFAMLLYPPGLEFICGYLGCLSAGVVAVPGVPPQGRSQNHRA
                     LTRMKRLIADADAKVILGREVIAALGTMAEHLPELAEITCVATEDIPDAAAGSWREP
                     DLTADSVAFLQYTSGSTSAPRGVMVTHGNLLDNERVITERMGHTPDTIAEYDHELFVS
                     WLPVYHDMGLIGPVLNTVYLGATATLFSPLHFLQQPQRWLTALSHYRPHTSGGPNFAY
                     ELCLKQATPDLLDGLDLSRWKVAFNGAEPVRAATLRRFTETFGTAGFRREALYPCYGL
                     AEATLMVTGGSVAAPPTLIEAPETGPHAGAADAAAVGSGRPGPGVTLVIADPERQQEL
                     PDGEVGEIWVLGASVAKGYWRNTLATRESFRATLKNREGRFLRTGDLGFLRDGELFVT
                     GRLKDLIVIDGRNHYPQDLELSAEMSHPALRPGCTATFSVDGGVEGEQPVIVAEAAPE
                     AADEFEKITDRIRSAIGEAHGLSVRDVVLVHPGTIPKTSSGKIQRRASRTAYLDGTLS
                     VVGAAATR"
     CDS             22404. .23084
                     /codon_start=1
                     /transl_table=11
                     /product="putative methyltransferase"
                     /protein_id="BAB69219.1"
                     /db_xref="GI:15824004"
                     /translation="MINSSNGTDYESMRFDRRGQAEAFDAIGDRYDEAFPHKEGQVSV
                     GEWLIGSLPPRSRVLDLGCGTGMPTARQLTDAGFDVVGVDLSGGMVELARAYVPGATF
                     HQLDVADLRPGGSRDLGRFDAVAAFFSLLMLPRPEIPLALLTVRHLLVPGGLFALAMV
                     EADVDDFSIPFVGHTIRVSGYLREDLHTVIETAGFEIVKEFSYTYAPSVADVPPEEQV
                     FLCCRRRG"
     CDS             23138. .25120
                     /codon_start=1
                     /transl_table=11
                     /product="putative membrane protein"
                     /protein_id="BAB69220.1"
                     /db_xref="GI:15824005"
                     /translation="MTEHHTPHEGPAQASRQSGGPLSGAGPVAPVTPMGPRVTRTPQP
                     PPAHGRHDGGRPGQPGKGPVTPSRADTGRHAPEGDAPEAEYAEGAVPDSTGPQTDRLR

  Query Match           63.6%;  Score 64.2;  DB 1;  Length 27522;
  Best Local Similarity  77.2%;  Pred. No. 0.0024;
  Matches   78;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           | ||| || ||| |||||  ||||||||||  |||||||  |||| |||||||||
Db   19146 GCTGCGCGGGCTGTCGCCGTCCGGGCGGTGCAGCCCGTTCTCCACCGAGGCGGACGGCGT 19087

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGC 101
           | | ||||||||||||||||   ||||  ||||||||||||
Db   19086 GGTGTGGGCCGAGGGCTGCGGTCTGGTCCTGCTCAAGCGGC 19046


RESULT 13
SAU421825
LOCUS       SAU421825              66808 bp    DNA     linear   BCT 16-MAR-2002
DEFINITION  Stigmatella aurantiaca ORF8, ORF7, ORF6, ORF5, ORF4, ORF3, ORF2,
            ORF1, stiA gene, stiB gene, stiC gene, stiD gene, stiE gene, stiF
            gene, stiG gene, stiH gene, stiJ gene, stiK gene, stiL gene and
            ORF9.
ACCESSION   AJ421825
VERSION     AJ421825.1  GI:19572309
KEYWORDS    acyl-CoA binding protein; cellulase; cytochrome P450 monooxygenase;
            kinase; methyl transferase; ORF1; ORF2; ORF3; ORF4; ORF5; ORF6;
            ORF7; ORF8; ORF9; polynucleotide adenyltransferase; potassium
            channel beta chain; ribosome binding factor; stiA gene; StiA
            protein; stiB gene; StiB protein; stiC gene; StiC protein; stiD
            gene; StiD protein; stiE gene; StiE protein; stiF gene; StiF
            protein; stiG gene; StiG protein; stiH gene; StiH protein; stiJ
            gene; StiJ protein; stiK gene; stiL gene.
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1
  AUTHORS   Gaitatzis,N., Silakowski,B., Kunze,B., Nordsiek,G., Blocker,H.,
            Hofle,G. and Muller,R.
  TITLE     The biosynthesis of the aromatic myxobacterial electron transport
            inhibitor stigmatellin is directed by a novel type of modular
            polyketide synthase
  JOURNAL   Online Publication
  REMARK    J. Biol. Chem., 10.1074/jbc.M111738200
REFERENCE   2  (bases 1 to 66808)
  AUTHORS   Muller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Muller R., MX, GBF, Mascheroderweg 1, 38124
            Braunschweig, GERMANY
FEATURES             Location/Qualifiers
     source          1. .66808
                     /organism="Stigmatella aurantiaca"
                     /strain="Sg a15"
                     /db_xref="taxon:41"
     CDS             complement(4978. .6231)
                     /note="ORF8"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P450 monooxygenase"
                     /protein_id="CAD19077.1"
                     /db_xref="GI:19572310"
                     /translation="MLRSVSASRPPSPANPVAAFDLARLDDAFYADPFPLYRAMRERD
                     PVHRMPDGSLFLTRWADLDCIYRDTRTFSSDKRAEFGAKYGDAPLFEHHTTSLVFNDP
                     PLHTRVRRLIVGALTPRALSTMEPGLRTLVDRLLDGLAVKGAADLIEDFAAAIPIEVI
                     GNLLDIPTEERGPLRGWSLAILGALEPRLTAEQEACGNEAVTEFLDYLRILVAQRRAR
                     PGDPATDVLTRLIQGESDGERLTETELLHQCVFLLNAGHETTTNLIGNALELLARFPD
                     ERARLLRAPALIPTAVEEVLRYESSNQLGNRRVAEDTEIGGVAVPTGTFLTLCIGAAN
                     RDPARFEDPEHFDVGRQPNRHLAFAGGAHTCAGMNLARIEARIALAAFLARFPDYALT
                     APPVRARRARFRGFTAMPTRLGSLR"
     CDS             6369. .7730
                     /note="ORF7"
                     /codon_start=1
                     /transl_table=11
                     /product="protein kinase"
                     /protein_id="CAD19078.1"
                     /db_xref="GI:19572311"
                     /translation="MEELISEALPDGSRVGPWRIEGRAGYGTYGAVYRAHRVGRICPR
                     SVALKLARYPNDLRFEREAELLERVRHANVPRLLGHGTWRAGPQGEAHPYVVMQWVEG
                     VRLYQWNKERLRTAREVLRVLAQVARALEATHMRGGLHRDVKGENVLVSPEGHASLMD
                     FGCGTWEGAAPLTGGILAPGTKAYRSPEALRFHWNHRNEADAHYQATPADDVYALGVT
                     AYRLCTGIYPPPATEPSGVGEDERETQEVKKPPSQVRPLVPALEQLILRMLRDNPKER
                     GRAGELAAEMEALAERAGAEADEPMNPNHSSVPSEITGRPVSVQREWGGWPVALVLVA
                     VGLLLWGSGSLNTGRRVAPASRMTEGGTGGVADAAVEALPVSSAAREPEPSGLNLDML
                     KEPLPGQRRPPCPRFQISIQGGCWLEVPQSSPPCGESYYEWKGACYSPVAAPPRPRTS
                     DPQ"
     CDS             complement(7849. .8577)
                     /note="ORF6"
                     /codon_start=1
                     /transl_table=11
                     /product="methyl transferase"
                     /protein_id="CAD19079.1"
                     /db_xref="GI:19572312"
                     /translation="MSQNIYDDPSFFESYSQLARSLKGLAGAPEWPALRALLPDVKDL
                     RVVDLGCGFGWFCRWARKNDAAHVLGLDLSEKMLTRAKAETSGDAITYQQADLERLVL
                     PEGAFDLAFSSLAIHYVEELEPLLATVHRGLTPGGWFVFSIEHPIFMASRQPEWMTGP
                     EGHRTWPVDSYQLEGKRTTHWLAEEVVKYHRTLGTTLNALLEQGFTLRHIEEWGPTDA
                     QLSAHPELAEERDRPMFLLVSAQR"
```

```
CDS             complement(8622. .8879)
                /note="ORF5"
                /codon_start=1
                /transl_table=11
                /product="acyl-CoA binding protein"
                /protein_id="CAD19080.1"
                /db_xref="GI:19572313"
                /translation="MALEDDFRSAQERVKALKTRPSNDRLLELYSLFKQATEGDVQGK
                RPGILDVKERAKYDAWVGRKGLGRESAMQQYVALVERLLRG"
CDS             8961. .10070
                /note="ORF4"
                /codon_start=1
                /transl_table=11
                /product="potassium channel beta chain"
                /protein_id="CAD19081.1"
                /db_xref="GI:19572314"
                /translation="MVYLLCRSRSQLRWILNGSAHRVRFAWGGHARPRRMHFRHLGRS
                GLMVSEISYGNWLTHGSQVEEAAAVACVRAALDEGITTFDTADVYAGTRAEEVLGRAL
                QGQRREGLEIFTKVYWPTGPGPNDRGLSRKHILESIHGSLRRLKTEYVDLYQAHRYDT
                ETPLEETMQAFSDVVRQGKAHYIGVSEWTADQLREGVKLSRQMGFQLVSSQPQYSMLW
                RVIEAEVVPASQELGVGQVVYSPIAQGVLTGKYRPGQKPPEGSRATDEKGGASFIKRF
                LNDDVLTRVQQLTPVAQEAGLSLAQLAVAWVLQNPNVSSAIIGASRPEQVKENVKAAG
                VKLSEEAMRHIDTLLGPIVERNPKKTLENSPRKRP"
CDS             10428. .10862
                /note="ORF3"
                /codon_start=1
                /transl_table=11
                /product="ribosome binding factor"
                /protein_id="CAD19082.1"
                /db_xref="GI:19572315"
                /translation="MTSFKQRRRRAPRRGTESSLSSFEAQSLPAHHLRVQSTLFQEVA
                RLFRGELSDPLLEGVSVTSFELTPDGRNARIGFTLPPEYAEGGSAPIEAALVRASGFI
                RAQLALGLNLKRVPHLRFVCVGVAPPGFSIPEDGSAEEGGES"
CDS             10859. .12016
                /note="ORF2"
                /codon_start=1
                /transl_table=11
                /protein_id="CAD19083.1"
                /db_xref="GI:19572316"
                /translation="MRPRRIEGAPGAVPILAWARAVPAGAVKQLQLLARQPYLAEHVA
                AMPDVHVAEGVAVGTVFATERHIVPGALGGDLGCGVSAVRFDFPAMSLGRVELTRLLA
                AFARAVPVGDAVHRGRGLSLPPGLDIGRLSTHRLQREWERLVPKHLGTLGGGNHFLEL
                DRDAEGALWLLLHTGSRGVGAAIAAHHLRVAQVSGEGSLPGLRTDTSEGAACLADTEL
                ACQFARANRDTLLQRAVEVLAGAIGQGPVLDSRVDVHHNHVAAESHFGRTLLVHRKGA
                VGLVAGQQGLIPGSMGTASYLVAGKGEPQAFCSCSHGAGRVMTRREARERIRPAALAH
                ALRRVVFDEGRVTALVEEAPTAYRDITEVLEDEAELVTPLRRLTPIAVLKG"
CDS             complement(12082. .13974)
                /note="ORF1"
                /codon_start=1
                /transl_table=11
                /product="cellulase"
                /protein_id="CAD19084.1"
                /db_xref="GI:19572317"
                /translation="MKNSLLLLMLAVLPLYGTAALARPGMANAEPGHFEGPENAGLTA
                AQCNPRVPLSTRGRYIVDRCGERFKLKSVNWFGASDQLEVVGGLDKQKLSDLVAGMKA
                MGFNSVRLPFSNNMLHPDDNKPAAQRCLEKYSVTCIDSTKNPELLNKTALEVYDAVVA
                ELTRQGLVVILNNHTTKSMWCCGWDGNGFWDSSQALQRWQDDWVMLAARYQGNKWVAG
                ADLRNEVRPDGLDSPNWGMRNQHDWHMAAQTMGNLLLRTNPDLLIVVEAVNWWGLLDG
                SRPQLKPVRQRPVALLRGDKLVYAVHNYGYTGPNQSGGSLGSGPKYSDMDKPTLYGTL
                DQEWGFALAANQAYTAPVWMSEFGVGYNEQAANSRAWFSNLADYLIEKDVDWAYWAIN
                AAKLQNSVQGEDETYGLWSYPDWNGVRSGDWRLGTGPLGRLLGADGRTGAVDATRFLP
                MTVLLKDGSSTYYVDNNSLDSDWHSGKAKISCPRGYRMVGVSANFTILCTNAGAVPAQ
                QGTGSYQTITQEVSPLRYPGDWANQSIKFECPTGFYAVGISTANPFFLELAGLLCEQN
                TGGIPLNSKSAKNFWGGDQRASSSGGDWSVSQYKGQCADNQYLIGLAHRRSGLADYPS
                VVLCSN"
gene            14345. .21466
                /gene="stiA"
CDS             14345. .21466
                /gene="stiA"
                /function="PKS"
                /codon_start=1
                /transl_table=11
                /product="StiA protein"
                /protein_id="CAD19085.1"
                /db_xref="GI:19572318"
                /translation="MTQLASRLGVGSESIDPRERFSRYGIDSRDTKSLLAELGALLKR
                PLSPTLAWTYPTVEALSRHLAGEADVSGGERADTLGVDQDEPIAIIGMAGRFPKAPDV
                ESFWSLLRDGVDAITEVPPERWDLDAFYSVGLKSPGKMNTRWGGFLDQVDQFDPLFFG
                ISPREAIQMDPQQRLALELSWEALESAGVAVDSLMDSRTGVFMGAMWSDYAKLTRGAL
                GLIGPHTATGQDSSIIAARVSYFLGLRGPSLTVNTACSSALVAVHLACQSLRNGEAAM
                ALAGGVNLLLSPESSVAMSQLGALSPDGRCKAFDARANGYVRGEGGGVVVLTPLSRAI
                ANGYPVLCVIRGSAVNNDGFSNGLTAPNPRAQEEVLRDAYARAGVSPDKVHYVETHGT
                GTVLGDPIEANALSAVLGVNRPAERPLILGSVKTNIGHLEAASGIAGLMKVVLSIQHR
                AIPPNLHFEQPNPYIRFEELRLLVPKSLHPWPNEGESALAGISSFGFGGTNCHLVLEA
                PSPGSAHIAPLAAETPEELSLLARQARDAVSVVGSQVELADLCFSVAARCREGSQRLA
                LIARSPAELAAQVDGFLAGEKRPGLLRGEVTSGGSQKPVFVFSGTGSQWPGMGRELLR
                TEPVFRGAIERCDRALRGLSGWSMLEELSAGEASSRLKECSGLQTAIFAIQVGLAALW
                RSWGIEPAGVVGHSIGEVAAAHVAGILNLEDAARVIHHLSRLMGRLSGQGAMIVVDLS
                AEEANRLLREWERHVVVAACNAPASTVLSGAREVIEELGAELERRGIAAHTVRIDTPS
                HSPLMDPFMGELAEVLAGLSPRVAAVPMVSTVTGAFLKAEECTAAYWVRNLREPVLFS
                QALTKLFEQDLRLFLELSPHPILARSVEQCLKGEGRQGVVLASMRRGENEGACMRESL
                GRLFTLGTSVAWQALYPAEAKRVELPMGLGTGSLDAELRGSARLLPLSAHTAEALEDA
                ARALQASIAKHPDVRLDDLCYSASVGKSVHAHRLAFVTDSKRALVEHLDDFLKGESRP
                GRIAGVCATGEPKKVVFVFPGQGSQWQGMGRQLLRQEPVFREAIEECDAAIRSCAGWS
                LIDALQADEPHSRLGSIEVIQPTLFAIEVALAALWRSWGVVPDAVVGHSMGETAAACV
                AGALRLTDAARVICERSQLLRRVAGQGGMALVELSLEQAREMLSRKDGRVSIGASNSP
                RSTVLSGDAIALEEILRELEAQGTFCRRVKVDVASHSPQVDPLKDDLLRVLSTIAPRA
                ATVPLYSTVTGEPIEGARMDASYWMRNLREPVLFGQAIERLAGDGHGVFVELSPHPIL
                LPSIDQTLRHKSHKGVVLPSLRREEPERAVLLESLGTLFTGGGRVDFRGLHPLGGHRV
                RLPHLPWQRQRFWLDMDVAAAGAASVGAAWHPGGGSEHPLLGHRLPSLAHAPACHVWE
                NKLSRSAVSYLDDHRVGGTVVLPGTSYIEMALAAAAKGSGKKTVSLRDVELREKMSFL
                ETASRTVQLSLIQEEAGHSSFRLHGRSTPDEPWTLHGGGTVHVEEGDGPPSADPVLWA
                AVETVRLQGKEVSASDAYEAMHAQGNEYGPLFRGIEQMWVRDAELVARLRVPEPLEQE
                LPEYRFHPAALDLGLQALIVLAAEHGCEGLALPTRLESIRAHASPGARMWCHVRLNPV
                REAASFSGEVLFLDESGQRLAELQGARFQILGEGSRPHDGGRLVDWLHEVEWQPAPVP
                VPTAEPSSQGRWVLFADGTGVAEGLAAQIAARGEPCIRVLPGMAFAEAEDGHFVVDPE
                QPEDVRKLFGALSRTSTIPLRGIVHLWSLDAPSPEEGGVDALSRARRLGSGSVLSLIQ
                GLNAAELAQKPRLWLVTRGVQAVGSAGPGAGAVAQSPLWGLGQVVAYEHAELRCTRVD
                LAPVSVPGEEQTLLQELLAPAGDDQVALRGRERHVPRLTRCRTDEVAGRSTDLRAGAI
                RSDGCYLITGGLGGLGLTIAKWLVARGARHLVLVGRRGASDAAEIALGQMRGAGAECL
                VMKADVANTSEVEELVGTIQRKMPPLRGIFHAAGLLDDGTVLQLTQARLRSVMAPKVE
                GAWNLHRATLGASLDFFVLFSSAASLVGIPGQGNYAAGNAFLDALAGLRKARGLACQS
                IQWGPWSEIGLAAEQDNRGERLALRGMGSLNPREGLEALEALLSPGEPLRERARIGVL
                RLDARRWCKSYPAIRSSPLFSVLLAGERNAKQVLERSGLRARLNEEPARTRRAALESF
                LREEAAQVLRLPPARIDLSRSLQTLGMDSLLSLELRNRLEANLELTLPATVLWNYPTL
                SALAPFLAERMGIQLNGVEEPPQAVESPSVSEEDGAAMLKMLEQFQELSNEDLLGMLS
                DDKEDEVSS"
gene            21463. .26220
                /gene="stiB"
CDS             21463. .26220
```

Query Match 63.2%; Score 63.8; DB 1; Length 66808;
Best Local Similarity 77.8%; Pred. No. 0.0022;
Matches 77; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```
Qy       1 GATGCACGCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTA 60
           ||| |    ||| || | || ||||||||||||||||||   |||  |||||| |||||
Db   59960 GATCCGGATGCTCTCACGCGACGGGCGGTGCAAGACGTTCGACGCCTCCGCGGATGGCTA 60019

Qy      61 CGCGCGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCG 99
           ||  ||| ||||||||||| |  || ||||||||||||||
Db   60020 CGGCCGGTCCGAGGGCTGTGGGGTCGTGGTGCTCAAGCG 60058
```

```
RESULT 14
AE007184/c
LOCUS       AE007184               14634 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 270 of 280 of the
            complete genome.
ACCESSION   AE007184 AE000516
VERSION     AE007184.1  GI:13883785
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
```

```
          Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1  (bases 1 to 14634)
  AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
          Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
          Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
          Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
          Gill,J., Mikula,A. and Bishai,W.
  TITLE   Whole genome comparison of Mycobacterium tuberculosis clinical and
          laboratory strains
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 14634)
  AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
          Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
          Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
          Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
          Gill,J., Mikula,A. and Bishai,W.
  TITLE   Direct Submission
  JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712
          Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .14634
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            120. .1136
                     /gene="MT3903"
     CDS             120. .1136
                     /gene="MT3903"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK48269.1"
                     /db_xref="GI:13883786"
                     /translation="MLGASACSPQKSPQPSSPRLPDNALITLGVAAGPPPTPSRVGIS
                     SVLKIGRDLYVIDCGLGSLNAFTNAGLQFDDLKAMFITHLHTDHIVDYYNFFLSGGFL
                     APPGRAPVLVYGPGPAGGLPPSEVGNPNPATVNPANPTPGLAAATEALHRAFAYTSNI
                     FIRDYGIDNVADLVKVTEIGLPPGSDYRNRAPKMSPFSVASDDNVSVTATLVSHYDVY
                     PAFGFRFDLKKSGVSVTFSGDTTKSDNLITLAQGTDILVHEAVFSLDTAYFGNAFPPN
                     YLVNSHTSAEQVGEVAAAAKPKQLILSHYAPDDLPDSQWLDKIKKNYSGMTTIARDGQ
                     VFAL"
     gene            1216. .2997
                     /gene="MT3904"
     CDS             1216. .2997
                     /gene="MT3904"
                     /note="similar to SP:P33224 GB:L20915 PID:457172
                     PID:457174 PID:537028; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="acyl-CoA dehydrogenase, putative"
                     /protein_id="AAK48270.1"
                     /db_xref="GI:13883787"
                     /translation="MPEYDLEAVDKLPFSTPEKAQRYQTENYRGAMGLNWYLTDPTLQ
                     FIMAYYLRPDELAFAEPHLTRIGELTGGPVTRWAEETDRNPPRLERYDRWGHDISRVV
                     LPESFIQSKRAVIEARQAVRDDAARAGVKPSLALFAADYLLNQADIGMACALATGGNM
                     VRSLVTAYAPPDVREFVLGKLNSGEWDGEAAQLLTERAGGSDLGALETTATRSGDVWL
                     LNGFKWFASNCAGEAFVVLAKPEGAPDSTRGVATFLVLRTRRDGSRNGVRIRRLKDKL
                     GTRSVASGEIEFVDAEAFLVGEPSADAGPSDGKGLTRMMELTNRLRLGTASFALGNA
                     RRALVESLCYAGQRRAFGGALIDKPLMRRKLAEMVVDVEAALAMVFDGFGAANHRQPR
                     CLPQRIAVPVTKLKTCRLGITVASDAIEIHGGNGYIETWPVARLLRDAQVNTIWEGPD
                     NILCLDVRRGIEQTRAHETLLARLRDAVSVSDDDDTTRLVSRRIEDLDAAITAWTKLD
                     RQLAEARLFPLAQFMGDVYAGALLTEQAAWERATGTDRKALVARLYARRYLADQGPL
                     RGIDADCDEALQRFDELVAGAFTAEQT"
     gene            3134. .4468
                     /gene="MT3905"
     CDS             3134. .4468
                     /gene="MT3905"
                     /note="similar to SP:Q10621 PID:1322437 GB:AL123456
                     GB:AL123456; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="IS1557, transposase"
                     /protein_id="AAK48271.1"
                     /db_xref="GI:13883788"
                     /translation="MRNVRLFRALLGVDKRTVIEDIEFEEDDAGDGARVIARVRPRSA
                     VLRRCGRCGRKASWYDRGAGLRQWRSLDWGTVEVFLEAEAPRVNCPTHGPTVVAVPWA
                     RHHAGHTYAFDDTVAWLAVACSKTAVCELMRIAWRTVGAIVARVWADTEKRIDRFANL
                     RRIGIDEISYKRHHRYLTVVVDHDSGRLVWAAPGHDKATLGLFFDALGAERAAQITHV
                     SADAADWIADVVTERCPDAIQCADPFHVVAWATEALDVERRRAWNDARAIARTEPKWG
                     RGRPGKNAAPRPGRERARRLKGARYALWKNPEDLTERQSAKLAWIAKTDPRLYRAYLL
                     KESLRHVFSVKGEEGKQALDRWISWAQRCRIPVFVELAARIKRHRVAIDAALDHGLSQ
                     GLIESTNTKIRLLTRIAFGFRSPQALIALAMLTLAGHRPTLPGRHNHPQISQ"
     gene            complement(4521. .6089)
                     /gene="MT3906"
     CDS             complement(4521. .6089)
                     /gene="MT3906"
                     /note="similar to SP:P53003; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="propionyl-CoA carboxylase, beta subunit"
                     /protein_id="AAK48272.1"
                     /db_xref="GI:13883789"
                     /translation="MTVTEPVLHTTAEKLAELRERLELAKEPGGEKAAAKRDKKGIPS
                     ARARIYELVDPGSFMEIGALCRTPGDPNALYGDGVVTGHGLINGRPVGVFSHDQTVFG
                     GTVGEMFGRKVARLMEWCAMVGCPIVGINDSGGARIQDAVTSLAWYAELGRRHELLSG
                     LVPQISIILGKCAGGAVYSPIQTDLVVAVRDQGYMFVTGPDVIKDVTGEDVSLDELGG
                     ADHQASYGNIHQVVESEAAAYQYVRDFLSFLPSNCFDKPPVVNPGLEPEITGHDLELD
                     SIVPDSDNMAYDMHEVLLRIFDDGDFLDVAAQAGQAIITGYARVDGRTVGVVANQPMH
                     MSGAIDNEASDKAARFIRFSDAFDIPLVFVVDTPGFLPGVEQEKNGIIKRGGRFLYAV
                     VEADVPKVTITIRKSYGGAYAVMGSKQLTADLNFAWPTARIAVIGADGAAQLLMKRFP
                     DPNAPEAQAIRKSFVENYNLNMAIPWIAAERGFIDAVIDPHETRLLLRKSMHLLRDKQ
                     LWWRVGRKHGLIPV"
     gene            complement(6086. .11287)
                     /gene="MT3907"
     CDS             complement(6086. .11287)
                     /gene="MT3907"
                     /note="similar to GB:U00023 PID:467207; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAK48273.1"
                     /db_xref="GI:13883790"
                     /translation="MADVAESQENAPAERAELTVPEMRQWLRNWVGKAVGKAPDSIDE
                     SVPMVELGLSSRDAVAMAADIEDLTGVTLSVAVAFAHPTIESLATRIIEGEPETDLAG
                     DDAEDWSRTGPAERVDIAIVGLSTRFPGEMNTPEQTWQALLEGRDGITDLPDGRWSEF
                     LEEPRLAARVAGARTRGGYLKDIKGFDSEFFAVAKTEADNIDPQQRMALELTWEALEH
                     ARIPASSLRGQAVGVYIGSSTNDYSFLAVSDPTVAHPYAITGTSSSIIANRVSYFYDF
                     HGPSVTIDTACSSSLVAIHQGVQALRNGEADVVVAGGVNALITPMVTLGFDEIGAVLA
                     PDGRIKSFSADADGYTRSEGGGMLVLKRVDDARRDGDAILAVIAGSAVNHDGRSNGLI
                     APNQDAQADVLRRAYKDAGIDPRTVDYIEAHGTGTILGDPIEAEALGRVVGRGRPADR
                     PALLGAVKTNVGHLESAAGAASMAKVVLALQHDKLPPSINFAGPSPYIDFDAMRLKMI
                     TTPTDWPRYGGYALAGVSSFGFGGANAHVVVREVLPRDVVEKEPEPEPEPKAAAEPAE
                     APTLAGHALRFDEFGNIITDSAVAEEPEPELPGVTEEALRLKEAALEELAAQEVTAPL
                     VPLAVSAFLTSRKKAAAAELADWMQSPEGQASSLESIGRSLSRRNHGRSRAVVLAHDH
                     DEAIKGLRAVAAGKQAPNVFSVDGPVTTGPVWVLAGFGAQHRKMGKSLYLRNEVFAAW
                     IEKVDALVQDELGYSVLELILDDAQDYGIETTQVTIFAIQIALGELLRHHGAKPAAVI
                     GQSLGEAASAYFAGGLSLRDATRAICSRSHLMGEGEAMLFGEYIRLMALVEYSADEIR
                     EVFSDFPDLEVCVYAAPTQTVIGGPPEQVDAILARAEAEGKFARKFATKGASHTSQMD
                     PLLGELTAELQGIKPTSPTCGIFSTVHEGRYIKPGGEPIHDVEYWKKGLRHSVYFTHG
                     IRNAVDSGHTTFLELAPNPVALMQVALTTADAGLHDAQLIPTLARKQDEVSSMVSTMA
                     QLYVYGHDLDIRTLFSRASGPQDYANIPPTRFKRKEHWLPAHFSGDGSTYMPGTHVAL
                     PDGRHVWEYAPRDGNVDLAALVRAAAAHVLPDAQLTAAEQRAVPGDGARLVTTMTRHP
                     GGASVQVHARIDESFTLVYDALVSRAGSESVLPTAVGAATAIAVADGAPVAPETPAED
                     ADAETLSDSLTTRYMPSGMTRWSPDSGETIAERLGLIVGSAMGYEPEDLPWEVPLIEL
                     GLDSLMAVRIKNRVEYDFDLPPIQLTAVRDANLYNVEKLIEYAVEHRDEVQQLHEHQK
                     TQTAEEIARAQAELLHGKVGKTEPVDSEAGVALPSPQNGEQPNPTGPALNVDVPPRDA
                     AERVTFATWAIVTGKSPGGIFNELPRLDDEAAAKIAQRLSERAEGPITAEDVLTSSNI
                     EALADKVRTYLEAGQIDGFVRTLRARPEAGGKVPVFVFHPAGGSTVVYEPLLGRLPAD
                     TPMYGFERVEGSIEERAQQYVPKLIEMQGDGPYVLVGWSLGGVLAYACAIGLRRLGKD
                     VRFVGLIDAVRAGEEIPQTKEEIRKRWDRYAAFAEKTFNVTIPAIPYEQLEELDDEGQ
                     VRFVLDAVSQSGVQIPAGIIEHQRTSYLDNRAIDTAQIQPYDGHVTLYMADRYHDDAI
                     MFEPRYAVRQPDGGWGEYVSDLEVVPIGGEHIQAIDEPIIAKVGEHMSRALGQIEADR
```

```
                TSEVGKQ"
     gene       complement(11294. .13183)
                /gene="MT3908"
     CDS        complement(11294. .13183)
                /gene="MT3908"
                /note="similar to PID:1658531; identified by sequence
                similarity; putative"
                /codon_start=1
                /transl_table=11
                /product="acyl-CoA synthase, putative"
                /protein_id="AAK48274.1"
                /db_xref="GI:13883791"
                /translation="MAYHNPFIVNGKIRFPANTNLVRHVEKWAKVRGDKLAYRFLDFS
                TERDGVARDILWSDFSARNRAVGARLQQVTQPGDRVAIXCPQNLDYLISFFGALYSGR
                IAVPLFDPAEPGHVGRLHAVLDDCAPSTILTTTDSAEGVRKFIRARSAKERPRVIAVD
                AVPTEVAATWQQPEANEETVAYLQYTSGSTRIPSGVQITHLNLPTNVVQVLNALEGQE
                GDRGVSWLPFFHDMGLITVLLASVLGHSFTFMTPAAFVRRPGRWIRELARKPGETGGT
                FSAAPNFAFEHAAVRGVPRDDEPPLDLSNVKGILNGSEPVSPASMRKFFEAFAPYGLK
                QTAVKPSYGLAEATLFVSTTPMDEVPTVIHVDRDELNNQRFVEVAADAPNAVAQVSAG
                KVGVSEWAVIVDADTASELPDGQIGEIWLHGNNLGTGYWGKEEESAQTFKNILKSRIS
                ESRAEGAPDDALWVRTGDYGTYFKDHLYIAGRIKDLVIIDGRNHYPQDLECTAQESTK
                ALRVGYAAAFSVPANQLPQTVFDDSHAGLKFDPEDTSEQLVIVGERAAGTHKLDHQPI
                VDDIRAAIAVGHGVTVRDVLLVSAGTIPRTSSGKIGRRACRAAYLDGSLRSGVGSPTV
                FATSD"
     gene       complement(13496. .14506)
                /gene="MT3909"
     CDS        complement(13496. .14506)
                /gene="MT3909"
                /note="identified by Glimmer2; putative"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AAK48275.1"
                /db_xref="GI:13883792"
                /translation="MAKNSRRKRHRILAWIAAGAMASVVALVIVAVVIMLRGAESPPS
                AVPPGVLPPGPTPAHPHKPRPAFQDASCPDVQMISVPGTWESSPQQNPLNPVQFPKAL
                LLKVTGPIAQQFAPARVQTYTVAYTAQFHNPLTTDNQMSYNDSRAEGTRAMVAAMTDM
                NNRCPLTSYVLIGFSQGAVIAGDVASDIGNGRGPVDEDLVLGVTLIADGRRQQGVGNQ
                VPPSPRGEGAEITLHEVPVLSGLGLTMTGPRPGGFGALDGRTNEICAQGDLICAAPAQ
                AFSPANLPTTLNTLAGGAGQPVHAMYATPEFWNSDGEPATEWTLNWAHQLIENAPHPK
                HR"
BASE COUNT     2365 a   4931 c   4768 g   2568 t      2 others
ORIGIN

  Query Match            62.0%;  Score 62.6;  DB 1;  Length 14634;
  Best Local Similarity   79.6%;  Pred. No. 0.0055;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy      8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
          | |||  |||||| ||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db  10294 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 10235

Qy     68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
          |||| ||| ||| | || ||||||||||||||||
Db  10234 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 10202


RESULT 15
MTV026/c
LOCUS       MTV026                 23740 bp    DNA     linear   BCT 03-AUG-2001
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.
ACCESSION   AL022076 AL123456
VERSION     AL022076.1  GI:3256026
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis H37Rv.
  ORGANISM  Mycobacterium tuberculosis H37Rv
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 23740)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
   PUBMED   9634230
REFERENCE   2  (bases 1 to 23740)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 26, 1998 this sequence version replaced gi:2961395.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES            Location/Qualifiers
     source         1. .23740
                    /organism="Mycobacterium tuberculosis H37Rv"
                    /strain="H37Rv"
                    /db_xref="taxon:83332"
     misc_feature   <1. .>23740
                    /note="fragment designated v026. Does not represent a
                    physical clone"
     RBS            60. .64
                    /note="possible RBS for Rv3797"
     gene           72. .1853
                    /gene="fadE35"
     CDS            72. .1853
                    /gene="fadE35"
                    /note="Rv3797, (MTV026.02), len: 593. fadE35, Probable
                    acyl Co-A dehydrogenase similar to E. coli AidB protein,
                    gp|L20915|ECOAIDB_3 E. coli (aidB) gene, comp (541 aa):
                    FASTA scores: opt: 455 z-score: 561.5 E(): 9.2e-24. 31.1%
                    identity in 514 aa overlap.Also similar to MTCI376_2 M.
                    tuberculosiscosmid  (542 aa),AidB, 31.9% identity in 458
                    aa overlap"
                    /codon_start=1
                    /transl_table=11
                    /product="fadE35"
                    /protein_id="CAA17861.1"
                    /db_xref="GI:2950416"
                    /db_xref="SPTREMBL:O53577"
                    /translation="MPEYDLEAVDKLPFSTPEKAQRYQTENYRGAMGLNWYLTDPTLQ
                    FIMAYYLRPDELAFAEPHLTRIGELTGGPVTRWAEETDRNPPRLERYDRWGHDISRVV
                    LPESFIQSKRAVIEARQAVRDDAARAGVKPSLALFAADYLLNQADIGMACALATGGNM
                    VRSLVTAYAPPDVREFVLGKLNSGEWDGEAAQLLTERAGGSDLGALETTATRSGDVWL
                    LNGFKWFASNCAGEAFVVLAKPEGAPDSTRGVATFLVLRTRRDGSRNGVRIRRLKDKL
                    GTRSVASGEIEFVDAEAFLLSGEPSADAGPSDGKGLTRMMELTNRLRLGTASFALGNA
                    RRALVESLCYAGQRRAFGGALIDKPLMRRKLAEMVVDVEAALAMVFDGFGAANHRQPR
                    CLPQRIAVPVTKLKTCRLGITVASDAIEIHGGNGYIETWPVARLLRDAQVNTIWEGPD
                    NILCLDVRRGIEQTRAHETLLARLRDAVSVSDDDDTTRLVSRRIEDLDAAITAWTKLD
                    RQLAEARLFPLAQFMGDVYAGALLTEQAAWERATRGTDRKALVARLYARRYLADQGPL
```

```
                RGIDADCDEALQRFDELVAGAFTAEQT"
gene            1980. .3314
                /gene="Rv3798"
CDS             1980. .3314
                /gene="Rv3798"
                /note="Rv3798, (MTV026.03), len: 444. Putative
                transposaseidentical to IS1557 MTCY373_32,
                sp|Q10621|Y02X_MYCTU HYPOTHETICAL 50.2 KD PROTEIN CY37
                (444 aa), 99.8% identity in 444aa overlap and similar to
                many, e.g.gp|U57649|DBU57649_1 Dibenzofuran-degrading
                bacterium (163 aa), FASTA scores: opt:  767 z-score:926.8
                E(): 0, 67.3% identity in 168 aa overlap, or to BPIS1001_1
                B.parapertussis insertion sequenc (406 aa), 24.9% identity
                in 402 aa overlap."
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv3798"
                /protein_id="CAA17862.1"
                /db_xref="GI:2950417"
                /db_xref="SWISS-PROT:Q10621"
                /translation="MRNVELFRALLGVDKRTVIEDIEFEEDDAGDGARVIARVRPRSA
                VLRRCGRCGRKASWYDRGAGLRQWRSLDWGTVEVFLEAEAPRVNCPTHGPTVVAVPWA
                RHHAGHTYAFDDTVAWLAVACSKTAVCELMRIAWRTVGAIVARVWADTEKRIDRFANL
                RRIGIDEISYKRHHRYLTVVVDHDSGRLVWAAPGHDKATLGLFFDALGAERAAQITHV
                SADAADWIADVVTERCPDAIQCADPFHVVAWATEALDVERRRAWNDARAIARTEPKWG
                RGRPGKNAAPRPGRERARELKGARYALWKNPEDLTERQSAKLAWIAKTDPRLYRAYLL
                KESLRHVFSVKGEEGKQALDRWISWAQRCRIPVFVELAARIKRHRVAIDAALDHGLSQ
                GLIESTNTKIRLLTRIAFGFRSPQALIALAMLTLAGHRPTLPGRHNHPQISQ"
repeat_unit     1980. .3311
                /gene="Rv3798"
                /note="IS1557-3rd copy"
gene            complement(3382. .4935)
                /gene="accD4"
CDS             complement(<3382. .4935)
                /gene="accD4"
                /note="Rv3799c, (MTV026.04c), accD4, len: 517. Probable
                Propionyl-CoA carboxylase; Beta chain, similar to many,
                e.g. sp|P53003|PCCB_SACER PROPIONYL-COA CARBOXYLASE BETA
                CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E():
                0; 50.7%identity in 519 aa overlap, Also similar to
                U00012_17 M. leprae cosmid B1308 (549 aa), 49.8%  identity
                in 514 aa overlap and to MTCY71_20 M. tuberculosis cosmid
                (548 aa), 48.9%identity in 515 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="accD4"
                /protein_id="CAA17863.1"
                /db_xref="GI:2950418"
                /db_xref="SPTREMBL:O53578"
                /translation="MTVTEPVLHTTAEKLAELRERLELAKEPGGEKAAAKRDKKGIPS
                ARARIYELVDPGSFMEIGALCRTPGDPNALYGDGVVTGHGLINGRPVGVFSHDQTVFG
                GTVGEMFGRKVARLMEWCAMVGCPIVGINDSGGARIQDAVTSLAWYAELGRRHELLSG
                LVPQISIILGKCAGGAVYSPIQTDLVVAVRDQGYMFVTGPDVIKDVTGEDVSLDELGG
                ADHQASYGNIHQVVESEADAYQYVRDFLSFLPSNCFDKPPVVNPGLEPEITGHDLELD
                SIVPDSDNMAYDMHEVLLRIFDDGDFLDVAAQAGQAIITGYARVDGRTVGVVANQPMH
                MSGAIDNEASDKAARFIRFSDAFDIPLVFVVDTPGFLPGVEQEKNGIIKRGGRFLYAV
                VEADVPKVTITIRKSYGGAYAVMGSKQLTADLNFAWPTARIAVIGADGAAQLLMKRFP
                DPNAPEAQAIRKSFVENYNLNMAIPWIAAERGFIDAVIDPHETRLLLRKSMHLLRDKQ
                LWWRVGRKHG"
gene            complement(4932. .10133)
                /gene="pks13"
CDS             complement(4932. .10133)
                /gene="pks13"
                /note="Rv3800c, (MTV026.05c), len: 1733. Probable
                polyketide synthase, similar to many. e.g.
                gp|M63676|SERERYAA_1 S.erythraea first ORF of eryA gene,
                involved in complex polyketide formation in erythromycin
                biosynthesis. FASTA scores: opt: 1329 z-score: 1411.0 E():
                0,  33.6% identity in 1232 aa overlap or to AF040570_9
                Amycolatopsis mediterranei rifamycin biosynthetic gene
                cluster (3413 aa),  34.3% identity in 1162 aa overlap.
                ALSO highly similar to mycobacterial putative polyketide
                synthase: MTCY338_19, U00023_3 M. leprae,
                MTCY22G10_2,MTCY19H9_2, U00023_6 M. leprae, contains
                PS00606 Beta-ketoacyl synthases active site"
                /codon_start=1
                /transl_table=11
                /product="pks13"
                /protein_id="CAA17864.1"
                /db_xref="GI:2950419"
                /db_xref="SPTREMBL:O53579"
                /translation="MADVAESQENAPAERAELTVPEMRQWLRNWVGKAVGKAPDSIDE
                SVPMVELGLSSRDAVAMAADIEDLTGVTLSVAVAFAHPTIESLATRIIEGEPETDLAG
                DDAEDWSRTGPAERVDIAIVGLSTRFPGEMNTPEQTWQALLEGRDGITDLPDGRWSEF
                LEEPRLAARVAGARTRGGYLKDIKGFDSEFFAVAKTEADNIDPQQRMALELTWEALEH
                ARIPASSLRGQAVGVYIGSSTNDYSFLAVSDPTVAHPYAITGTSSSIIANRVSYFYDF
                HGPSVTIDTACSSSLVAIHQGVQALRNGEADVVVAGGVNALITPMVTLGFDEIGAVLA
                PDGRIKSFSADADGYTRSEGGGMLVLKRVDDARRDGDAILAVIAGSAVNHDGRSNGLI
                APNQDAQADVLRRAYKDAGIDPRTVDYIEAHGTGTILGDPIEAEALGRVVGRGRPADR
                PALLGAVKTNVGHLESAAGAASMAKVVLALQHDKLPPSINFAGPSPYIDFDAMRLKMI
                TTPTDWPRYGGYALAGVSSFGFGGANAHVVVREVLPRDVVEKEPEPEPEPKAAAEPAE
                APTLAGHALRFDEFGNIITDSAVAEEPEPELPGVTEEALRLKEAALEELAAQEVTAPL
                VPLAVSAFLTSRKKAAAAELADWMQSPEGQASSLESIGRSLSRRNHGRSRAVVLAHDH
                DEAIKGLRAVAAGKQAPNVFSVDGPVTTGPVWVLAGFGAQHRKMGKSLYLRNEVFAAW
                IEKVDALVQDELGYSVLELILDDAQDYGIETTQVTIFAIQIALGELLRHHGAKPAAVI
                GQSLGEAASAYFAGGLSLRDATRAICSRSHLMGEGEAMLFGEYIRLMALVEYSADEIR
                EVFSDFPDLEVCVYAAPTQTVIGGPPEQVDAILARAEAEGKFARKFATKGASHTSQMD
                PLLGELTAELQGIKPTSPTCGIFSTVHEGRYIKPGGEPIHDVEYWKKGLRHSVYFTHG
                IRNAVDSGHTTFLELAPNPVALMQVALTTADAGLHDAQLIPTLARKQDEVSSMVSTMA
                QLYVYGHDLDIRTLFSRASGPQDYANIPPTRFKRKEHWLPAHFSGDGSTYMPGTHVAL
                PDGRHVWEYAPRDGNVDLAALVRAAAAHVLPDAQLTAAEQRAVPGDGARLVTTMTRHP
                GGASVQVHARIDESFTLVYDALVSRAGSESVLPTAVGAATAIAVADGAPVAPETPAED
                ADAETLSDSLTTRYMPSGMTRWSPDSGETIAERLGLIVGSAMGYEPEDLPWEVPLIEL
                GLDSLMAVRIKNRVEYDFDLPPIQLTAVRDANLYNVEKLIEYAVEHRDEVQQLHEHQK
                TQTAEEIARAQAELLHGKVGKTEPVDSEAGVALPSPQNGEQPNPTGPALNVDVPPRDA
                AERVTFATWAIVTGKSPGGIFNELPRLDDEAAAKIAQRLSERAEGPITAEDVLTSSNI
                EALADKVRTYLEAGQIDGFVRTLRARPEAGGKVPVFVFHPAGGSTVVYEPLLGRLPAD
                TPMYGFERVEGSIEERAQQYVPKLIEMQGDGPYVLVGWSLGGVLAYACAIGLRRLGKD
                VRFVGLIDAVRAGEEIPQTKEEIRKRWDRYAAFAEKTFNVTIPAIPYEQLEELDDEGQ
                VRFVLDAVSQSGVQIPAGIIEHQRTSYLDNRAIDTAQIQPYDGHVTLYMADRYHDDAI
                MFEPRYAVRQPDGGWGEYVSDLEVVPIGGEHIQAIDEPIIAKVGEHMSRALGQIEADR
                TSEVGKQ"
    RBS         complement(4945. .4948)

  Query Match             62.0%;  Score 62.6;  DB 1;  Length 23740;
  Best Local Similarity   79.6%;  Pred. No. 0.0049;
  Matches   74;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 GCGCTTTCGCCCGGCGGGCGGTGCAAGACGTTCTCGGCCGACGCGGACGGCTACGCGCGG 67
           | |||  |||||| ||| |||  |||| ||||||| |||||||| ||||||||| | ||
Db    9140 GTGCTGGCGCCCGACGGCCGGATCAAGTCGTTCTCAGCCGACGCCGACGGCTACACCCGC 9081

Qy      68 GCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGG 100
            |||| ||| ||| | || ||||||||||||||
Db    9080 TCCGAAGGCGGCGGCATGCTGGTGCTCAAGCGG 9048


Search completed: November  5, 2002, 12:16:41
Job time : 602.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590
Perfect score: 101
Sequence: 1 ctcgacgtgcgacgaggctt.....ctgggtatgccgctgtccgc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	2	99.4	98.4	68750	3	US-09-335-409-1	Sequence 1, Appli
	3	99.4	98.4	68750	4	US-09-568-102-1	Sequence 1, Appli
	4	99.4	98.4	68750	4	US-09-567-969-1	Sequence 1, Appli
	5	99.4	98.4	68750	4	US-09-568-480-1	Sequence 1, Appli
	6	99.4	98.4	68750	4	US-09-568-486-1	Sequence 1, Appli
	7	99.4	98.4	68750	4	US-09-568-472-1	Sequence 1, Appli
	8	99.4	98.4	68750	4	US-09-567-899-1	Sequence 1, Appli
	9	42.4	42.0	13842	4	US-09-105-537-30	Sequence 30, Appl
	10	42.4	42.0	13987	2	US-08-804-227C-13	Sequence 13, Appl
	11	42.4	42.0	36778	4	US-09-105-537-5	Sequence 5, Appli
	12	42.4	42.0	38506	3	US-09-320-878-19	Sequence 19, Appl
	13	42.4	42.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
	14	42.4	42.0	44377	2	US-08-804-227C-7	Sequence 7, Appli
	15	42.4	42.0	44377	2	US-08-804-198-1	Sequence 1, Appli
	16	40.8	40.4	11220	4	US-09-105-537-32	Sequence 32, Appl
	17	39.8	39.4	4689	4	US-09-105-537-34	Sequence 34, Appl
	18	39.2	38.8	1681	4	US-09-434-288-7	Sequence 7, Appli
	19	39.2	38.8	20235	1	US-07-642-734C-3	Sequence 3, Appli
	20	39.2	38.8	20235	3	US-08-439-009A-3	Sequence 3, Appli
	21	38.2	37.8	8460	1	US-08-469-005A-9	Sequence 9, Appli
	22	38.2	37.8	8519	4	US-09-261-907-1	Sequence 1, Appli
	23	37.6	37.2	15872	4	US-09-105-537-1	Sequence 1, Appli
	24	37.6	37.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
	25	37.6	37.2	80161	4	US-09-370-700-1	Sequence 1, Appli
	26	37.4	37.0	33529	4	US-09-144-085-3	Sequence 3, Appli
	27	36.4	36.0	11219	1	US-07-642-734C-1	Sequence 1, Appli
	28	36.4	36.0	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	29	34.2	33.9	748	3	US-09-154-083-1	Sequence 1, Appli
c	30	34.2	33.9	753	3	US-09-154-083-28	Sequence 28, Appl
	31	33.4	33.1	50937	4	US-09-428-517-1	Sequence 1, Appli
c	32	33	32.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	33	32.8	32.5	444	4	US-09-060-756-658	Sequence 658, App
	34	32.2	31.9	28958	1	US-08-258-261B-6	Sequence 6, Appli
	35	32.2	31.9	28958	1	US-08-456-837-6	Sequence 6, Appli
	36	32.2	31.9	28958	1	US-08-457-342-6	Sequence 6, Appli
	37	32.2	31.9	28958	1	US-08-457-646A-6	Sequence 6, Appli
	38	32.2	31.9	28958	1	US-08-458-076A-6	Sequence 6, Appli
	39	32.2	31.9	28958	1	US-08-764-233A-4	Sequence 4, Appli
	40	32.2	31.9	28958	1	US-08-457-335A-6	Sequence 6, Appli
	41	32.2	31.9	28958	1	US-08-729-214-6	Sequence 6, Appli
	42	32.2	31.9	28958	3	US-09-028-934-6	Sequence 6, Appli
	43	32.2	31.9	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	44	31.2	30.9	1434	4	US-09-434-288-3	Sequence 3, Appli
	45	30.6	30.3	4041	4	US-09-105-537-36	Sequence 36, Appl

ALIGNMENTS

RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 3.8e-22;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20490 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 20549

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||||
Db   20550 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 20590
```

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             98.4%;  Score 99.4;  DB 3;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCA GGCCTCGACTCCCTGATGGCTGTGGAGATC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   26102 CTCGACGTGCGACGAGGCTTCGCCGAGCA GGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTAT CCGCTGTCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   26162 CGCAAACGGCTTCAGGGTGAGCTGGGTAT CCGCTGTCGGC 26202


RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 4
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 1.2e-21;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 9
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-30

  Query Match             42.0%;  Score 42.4;  DB 4;  Length 13842;
  Best Local Similarity   64.0%;  Pred. No. 0.00027;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  13286 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13345

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |     ||  |  | ||  ||||
```

```
Db  13346 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13385


RESULT 10
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  13987 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..13987
US-08-804-227C-13

  Query Match             42.0%;  Score 42.4;  DB 2;  Length 13987;
  Best Local Similarity   64.0%;  Pred. No. 0.00027;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||      ||||  ||||   |||    || ||||||||||||  ||| ||||||  | |
Db  13551 TCGAGCCCGAACGACCCTTCCGGGAGATCGGCGTCGACTCCCTGGCGGCGGTGGAGCTGC 13610

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           ||||  |||||     ||  | | ||  ||| | || || |
Db  13611 GCAACCGGCTCAGCCGTCTGGTCGGCCTGCGGTTGCCGAC 13650


RESULT 11
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             42.0%;  Score 42.4;  DB 4;  Length 36778;
  Best Local Similarity   64.0%;  Pred. No. 0.00032;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  15027 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 15086

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           ||||  |||||  |    |      ||  |  | ||  ||||
Db  15087 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 15126


RESULT 12
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             42.0%;  Score 42.4;  DB 3;  Length 38506;
  Best Local Similarity   64.0%;  Pred. No. 0.00032;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  13169 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13228

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           ||||  |||||  |    |      ||  |  | ||  ||||
Db  13229 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13268


RESULT 13
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             42.0%;  Score 42.4;  DB 2;  Length 43280;
  Best Local Similarity   64.0%;  Pred. No. 0.00033;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||| || ||  |||||   ||   |||| ||||||||||  |  || || ||  | |
Db   19512 TCGACGTCCGCCGGAGCTTCAAGGAAGCGGGCTTCGACTCCCTCACCGCCGTCGAACTGC 19571

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| ||||| |  |  |     ||  ||  |||| | ||
Db   19572 GCAACCGGCTGCGCGCCGCCACCGGGCTGAAGCTGCCCGC 19611


RESULT 14
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             42.0%;  Score 42.4;  DB 2;  Length 44377;
  Best Local Similarity   64.0%;  Pred. No. 0.00033;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||      ||||  ||||   |||   ||| ||||||||||  ||| ||||||  | |
Db   13566 TCGAGCCCGAACGACCCTTCCGGGAGATCGGCGTCGACTCCCTGGCGGCGGTGGAGCTGC 13625

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||     ||  | | ||  ||| | || || |
Db   13626 GCAACCGGCTCAGCCGTCTGGTCGGCCTGCGGTTGCCGAC 13665


RESULT 15
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
```

```
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1118
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             42.0%;  Score 42.4;  DB 2;  Length 44377;
  Best Local Similarity   64.0%;  Pred. No. 0.00033;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGC GCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||      ||||  ||||   |||   ||| ||||||||||  ||| ||||||  | |
Db   13566 TCGAGCCCGAACGACCCTTCCGGGAGATCG GCGTCGACTCCCTGGCGGCGGTGGAGCTGC 13625

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGC CGCTGTCGGC 101
           |||| |||||     ||  | | ||  ||| | || || |
Db   13626 GCAACCGGCTCAGCCGTCTGGTCGGCCTGC GGTTGCCGAC 13665


Search completed:  November  5, 2002, 14:07:34
Job time :  171.737 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590
Perfect score: 101
Sequence: 1 ctcgacgtgcgacgaggctt.....ctgggtatgccgctgtcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	42.4	42.0	13842	10	US-09-861-289-30	Sequence 30, Appl
	2	42.4	42.0	36778	10	US-09-861-289-5	Sequence 5, Appli
	3	40.8	40.4	11220	10	US-09-861-289-32	Sequence 32, Appl
	4	39.8	39.4	4689	10	US-09-861-289-34	Sequence 34, Appl
	5	38.2	37.8	3209	10	US-09-925-301-474	Sequence 474, App
	6	37.6	37.2	15872	10	US-09-861-289-1	Sequence 1, Appli
	7	30.6	30.3	4041	10	US-09-861-289-36	Sequence 36, Appl
c	8	29.4	29.1	1083	10	US-09-815-242-7578	Sequence 7578, Ap
	9	29.4	29.1	3189	12	US-10-044-090-333	Sequence 333, App
	10	29.4	29.1	3299	10	US-09-800-729-68	Sequence 68, Appl
	11	29.4	29.1	6065	10	US-09-800-729-35	Sequence 35, Appl
	12	28.8	28.5	284	10	US-09-294-093B-4947	Sequence 4947, Ap
	13	28.4	28.1	1830	10	US-09-731-872-241	Sequence 241, App
	14	28.4	28.1	1831	10	US-09-731-872-240	Sequence 240, App
	15	28.4	28.1	2492	10	US-09-925-301-593	Sequence 593, App
c	16	27	26.7	884	10	US-09-833-381-1048	Sequence 1048, Ap
	17	27	26.7	3132	10	US-09-815-242-4139	Sequence 4139, Ap
c	18	27	26.7	167343	10	US-09-962-436-281	Sequence 281, App
c	19	27	26.7	167343	10	US-09-964-824A-273	Sequence 273, App
c	20	26.8	26.5	1161	10	US-09-808-387-5	Sequence 5, Appli
c	21	26.8	26.5	1674	10	US-09-808-387-3	Sequence 3, Appli
c	22	26.8	26.5	1725	10	US-09-808-387-1	Sequence 1, Appli
	23	26.4	26.1	2298	10	US-09-866-582-21	Sequence 21, Appl
	24	26	25.7	272	10	US-09-878-574-5740	Sequence 5740, Ap
c	25	26	25.7	577	10	US-09-864-761-9695	Sequence 9695, Ap
c	26	26	25.7	1023	10	US-09-814-777A-16	Sequence 16, Appl
c	27	26	25.7	1155	10	US-09-814-777A-21	Sequence 21, Appl
c	28	26	25.7	1421	10	US-09-814-777A-14	Sequence 14, Appl
c	29	26	25.7	1730	10	US-09-814-777A-19	Sequence 19, Appl
c	30	26	25.7	1919	10	US-09-814-777A-17	Sequence 17, Appl
	31	25.8	25.5	257	10	US-09-923-876-3033	Sequence 3033, Ap
	32	25.8	25.5	3186	10	US-09-880-107-1659	Sequence 1659, Ap
c	33	25.8	25.5	4358	10	US-09-917-800A-1343	Sequence 1343, Ap
	34	25.8	25.5	13029	10	US-09-815-242-4052	Sequence 4052, Ap
	35	25.6	25.3	1888	10	US-09-925-300-722	Sequence 722, App
c	36	25.6	25.3	2001	10	US-09-768-877-7	Sequence 7, Appli
c	37	25.6	25.3	2516	10	US-09-768-877-11	Sequence 11, Appl
	38	25.6	25.3	3177	10	US-09-898-570-15	Sequence 15, Appl
	39	25.6	25.3	35871	10	US-09-956-335-2	Sequence 2, Appli
	40	25.6	25.3	35935	10	US-09-725-720-43	Sequence 43, Appl
	41	25.6	25.3	35935	10	US-09-782-378A-4	Sequence 4, Appli
	42	25.6	25.3	35935	10	US-09-782-378A-5	Sequence 5, Appli
	43	25.6	25.3	35978	10	US-09-956-335-1	Sequence 1, Appli
c	44	25.4	25.1	147	10	US-09-864-761-24022	Sequence 24022, A
	45	25.4	25.1	487	10	US-09-833-381-1125	Sequence 1125, Ap

ALIGNMENTS

RESULT 1
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 42.0%; Score 42.4; DB 10; Length 13842;
Best Local Similarity 64.0%; Pred. No. 0.00016;
Matches 64; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

```
Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db   13286 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13345

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
             |||| |||||  |    |     ||  |  | ||  ||||
Db   13346 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13385
```

RESULT 2
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/851,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             42.0%;  Score 42.4;  DB 10;  Length 36778;
  Best Local Similarity   64.0%;  Pred  No. 0.00019;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db   15027 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 15086

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||  |    |     ||  |  | ||  ||||
Db   15087 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 15126


RESULT 3
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             40.4%;  Score 40.8;  DB 10;  Length 11220;
  Best Local Similarity   63.0%;  Pred. No. 0.00049;
  Matches   63;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          |||||  | || | ||||   || |   |  ||||||| |||| ||| || ||| |||
Db    4400 TCGACACGGGGCGGGCCTTCCGTGACCTCGGATTCGACTCGCTGACGGCGGTCGAGCTCC 4459

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| || ||  ||   |     ||  || | ||  ||||
Db    4460 GCAACCGCCTCAAGAACGCCACCGGCCTGCCCTCCCGGC 4499


RESULT 4
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             39.4%;  Score 39.8;  DB 10;  Length 4689;
  Best Local Similarity   64.8%;  Pred. No. 0.00086;
  Matches   59;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      11 GACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGGC 70
           | || | |||| ||||||  ||| |||||||| |||| ||| ||| || ||||||| | ||
Db    4283 GCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGCTCCGCAACCAGC 4342

Qy      71 TTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |       | | | || |    |||  | ||
Db    4343 TCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373


RESULT 5
US-09-925-301-474
; Sequence 474, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 474
;   LENGTH: 3209
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (427)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-474

  Query Match             37.8%;  Score 38.2;  DB 10;  Length 3209;
  Best Local Similarity   61.6%;  Pred. No. 0.0025;
  Matches   61;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           || || ||    |     | || || | |||||| ||||| || |||   |||||| | |
Db    1263 TCAACCTGGACAGCTCACTGGCGGACCTGGGCCTGGACTCGCTCATGAGCGTGGAGGTGC 1322

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGG 100
           || |   |||  || |||||||     ||  |||||| |
Db    1323 GCCAGACGCTGGAGCGTGAGCTCAACCTGGTGCTGTCCG 1361


RESULT 6
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             37.2%;  Score 37.6;  DB 10;  Length 15872;
  Best Local Similarity   61.0%;  Pred. No. 0.0051;
  Matches   61;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           |||  | |   ||   ||||   || |  ||| ||||||||||||   |  |||||| |||
Db    8253 TCGGGGCGGAGCGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCC 8312

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||  | |  |     ||  ||| |||  | ||
Db    8313 GCAACCGGCTGAACGCCGCCACCGGCCTGCGGCTCCCCGC 8352


RESULT 7
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-36

  Query Match             30.3%;  Score 30.6;  DB 10;  Length 4041;
  Best Local Similarity   62.3%;  Pred. No. 0.62;
  Matches   48;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      25 GAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGGCTTCAGGGTGAGCTG 84
           |||   |||·|||||||| ||||  || || ||  ||||||| ||  |  |  |  |
Db    2923 GAGATCGGCTTCGACTCGCTGACCGCCGTCGACTTCCGCAACCGCGTCAACCGGCTGACC 2982

Qy      85 GGTATGCCGCTGTCGGC 101
           ||| | | |||| || |
Db    2983 GGTCTCCAGCTGCCGCC 2999


RESULT 8
US-09-815-242-7578/c
; Sequence 7578, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7578
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1083)
US-09-815-242-7578

  Query Match             29.1%;  Score 29.4;  DB 10;  Length 1083;
  Best Local Similarity   63.4%;  Pred. No. 1.2;
  Matches   45;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||| |   || || ||| ||||||||||| ||||  ||| | | |||||  | |     |
Db      92 TCGGCAATCGTCGCGGCATCGCCGAGCAGCGCCTGCACTTCTTCATGGCGCTCGTACAAC 33

Qy      62 GCAAACGGCTT 72
           ||   | | ||
Db      32 GCTTCCAGTTT 22


RESULT 9
US-10-044-090-333
; Sequence 333, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 333
;   LENGTH: 3189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 2960079CB1
US-10-044-090-333

  Query Match             29.1%;  Score 29.4;  DB 12;  Length 3189;
  Best Local Similarity   58.6%;  Pred. No. 1.4;
```

```
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      10 CGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGG 69
           | | | |||   |  ||||||| ||   ||| ||    ||| | |||| || |||| | |
Db     179 CCAAGCGGCGGGGAAGAGCAGGCCCAGTACTGCCGCGCGGCGGAGGAGCTCAGCAAGCTG 238

Qy      70 CTTCAGGGTGAGCTGGGTATGCCGCTG 96
           |  |  |  |   | |||   ||||||
Db     239 CGCCGCGCCGCAGTCGGTCGTCCGCTG 265


RESULT 10
US-09-800-729-68
; Sequence 68, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
;  APPLICANT:  Ni et al.
;  TITLE OF INVENTION:  32 Human secreted proteins
;  FILE REFERENCE:  PZ044P1
;  CURRENT APPLICATION NUMBER: US/09/800,729
;  CURRENT FILING DATE:  2001-03-08
;  PRIOR APPLICATION NUMBER:  PCT/US00/26013
;  PRIOR FILING DATE:  2000-09-22
;  PRIOR APPLICATION NUMBER:  60/155,709
;  PRIOR FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  217
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 68
;   LENGTH: 3299
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-800-729-68

  Query Match             29.1%;  Score 29.4;  DB 10;  Length 3299;
  Best Local Similarity   58.6%;  Pred. No. 1.4;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      10 CGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGG 69
           | | | |||   |  ||||||| ||   ||| ||    ||| | |||| || |||| | |
Db     249 CCAAGCGGCGGGGAAGAGCAGGCCCAGTACTGCCGCGCGGCGGAGGAGCTCAGCAAGCTG 308

Qy      70 CTTCAGGGTGAGCTGGGTATGCCGCTG 96
           |  |  |  |   | |||   ||||||
Db     309 CGCCGCGCCGCAGTCGGTCGTCCGCTG 335


RESULT 11
US-09-800-729-35
; Sequence 35, Application US/09800729
; Patent No. US20020068319A1
; GENERAL INFORMATION:
;  APPLICANT:  Ni et al.
;  TITLE OF INVENTION:  32 Human secreted proteins
;  FILE REFERENCE:  PZ044P1
;  CURRENT APPLICATION NUMBER: US/09/800,729
;  CURRENT FILING DATE:  2001-03-08
;  PRIOR APPLICATION NUMBER:  PCT/US00/26013
;  PRIOR FILING DATE:  2000-09-22
;  PRIOR APPLICATION NUMBER:  60/155,709
;  PRIOR FILING DATE:  1999-09-24
;  NUMBER OF SEQ ID NOS:  217
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 35
;   LENGTH: 6065
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (6035)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (6037)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (6038)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-800-729-35

  Query Match             29.1%;  Score 29.4;  DB 10;  Length 6065;
  Best Local Similarity   58.6%;  Pred. No. 1.6;
  Matches   51;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      10 CGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGG 69
           | | | |||   |  ||||||| ||   ||| ||    ||| | |||| || |||| | |
Db     249 CCAAGCGGCGGGGAAGAGCAGGCCCAGTACTGCCGCGCGGCGGAGGAGCTCAGCAAGCTG 308

Qy      70 CTTCAGGGTGAGCTGGGTATGCCGCTG 96
           |  |  |  |   | |||   ||||||
Db     309 CGCCGCGCCGCAGTCGGTCGTCCGCTG 335


RESULT 12
US-09-294-093B-4947
; Sequence 4947, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
;  APPLICANT:       Lalgudi, Raghunath, V.
;  APPLICANT:  Ito, Laura, Y.
;  APPLICANT:  Sherman, Bradley, K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
;  FILE REFERENCE:  PL-0009 US
;  CURRENT APPLICATION NUMBER: US/09/294,093B
;  CURRENT FILING DATE:  1999-04-16
;  PRIOR APPLICATION NUMBER:  60/082,567
;  PRIOR FILING DATE:  April 21, 1998
;  NUMBER OF SEQ ID NOS:      6207
;  SOFTWARE:       PERL Program
; SEQ ID NO 4947
;   LENGTH: 284
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20010051335A1 700355442H1
US-09-294-093B-4947

  Query Match             28.5%;  Score 28.8;  DB 10;  Length 284;
  Best Local Similarity   60.0%;  Pred. No. 1.4;
  Matches   48;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||   |  |||       |||  | ||| ||| ||       | ||||||||||
Db      34 TCGACGACTGGAGAGAGGCACACGAAGAAGGCGTCGTCTTTTGCCCGACTGTGGAGATCA 93

Qy      62 GCAAACGGCTTCAGGGTGAG 81
           ||||||||  | | | |||
Db      94 TCAAACGGCGACGGCGGGAG 113


RESULT 13
US-09-731-872-241
; Sequence 241, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Jobert, Severin
;  TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
;  FILE REFERENCE: 78.US3.REG
;  CURRENT APPLICATION NUMBER: US/09/731,872
;  CURRENT FILING DATE:  2000-12-07
;  PRIOR APPLICATION NUMBER: US 60/169,629
;  PRIOR FILING DATE: 1999-12-08
;  PRIOR APPLICATION NUMBER: US 60/187,470
;  PRIOR FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 482
```

```
;  SOFTWARE: Patent.pm
; SEQ ID NO 241
;   LENGTH: 1830
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 78..608
US-09-731-872-241

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 1830;
  Best Local Similarity   58.1%;  Pred. No. 2.6;
  Matches   50;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      16 GGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGGCTTCAG 75
           | | |||  |||| ||  |  |     |||  |||||||| |  ||  |    || ||
Db     683 GTCCTCGGGGAGCTGGTGCGGGCGAAGCTGCCGGCTGTGGGGGCCCTGATGGAGCGTCTC 742

Qy      76 GGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||| | | || ||||  | |
Db     743 GGTGTGCTGTGGACGCTGCTGGTGTC 768


RESULT 14
US-09-731-872-240
; Sequence 240, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, Jean Baptiste
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Jobert, Severin
;  TITLE OF INVENTION: FULL-LENGTH HUMAN cDNAs ENCODING POTENTIALLY SECRETED PROTEINS
;  FILE REFERENCE: 78.US3.REG
;  CURRENT APPLICATION NUMBER: US/09/731,872
;  CURRENT FILING DATE:  2000-12-07
;  PRIOR APPLICATION NUMBER: US 60/169,629
;  PRIOR FILING DATE: 1999-12-08
;  PRIOR APPLICATION NUMBER: US 60/187,470
;  PRIOR FILING DATE: 2000-03-06
;  NUMBER OF SEQ ID NOS: 482
;  SOFTWARE: Patent.pm
; SEQ ID NO 240
;   LENGTH: 1831
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 78..590
US-09-731-872-240

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 1831;
  Best Local Similarity   58.1%;  Pred. No. 2.6;
  Matches   50;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      16 GGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGGCTTCAG 75
           | | |||  |||| ||  |  |     |||  |||||||| |  ||  |    || ||
Db     684 GTCCTCGGGGAGCTGGTGCGGGCGAAGCTGCCGGCTGTGGGGGCCCTGATGGAGCGTCTC 743

Qy      76 GGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||| | | || ||||  | |
Db     744 GGTGTGCTGTGGACGCTGCTGGTGTC 769


RESULT 15
US-09-925-301-593
; Sequence 593, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA106
;  CURRENT APPLICATION NUMBER: US/09/925,301
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05882
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 1694
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 593
;   LENGTH: 2492
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2113)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc_feature
;   LOCATION: (2452)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-593

  Query Match             28.1%;  Score 28.4;  DB 10;  Length 2492;
  Best Local Similarity   58.1%;  Pred. No. 2.7;
  Matches   50;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      16 GGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCCGCAAACGGCTTCAG 75
           | | |||  |||| ||  |  |     |||  |||||||| |  ||  |    || ||
Db     671 GTCCTCGGGGAGCTGGTGCGGGCGAAGCTGCCGGCTGTGGGGGCCCTGATGGAGCGTCTC 730

Qy      76 GGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||| | | || ||||  | |
Db     731 GGTGTGCTGTGGACGCTGCTGGTGTC 756


Search completed: November  5, 2002, 22:59:50
Job time : 97.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590
Perfect score: 101
Sequence: 1 ctcgacgtgcgacgaggctt.....ctgggtatgccgctgtcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	71989	21	AAA29349	Sorangium cellulos
2	99.4	98.4	68750	21	AAZ55887	Sorangium cellulos
3	52	51.5	20394	22	AAF24892	Pimaricin biosynth
4	44	43.6	30690	21	AAA92301	S. avermitilis ave
5	44	43.6	30690	22	AAH79277	Streptomyces averm
6	44	43.6	65140	22	AAD17184	Streptomyces nours
7	44	43.6	125401	22	AAD17186	Streptomyces nours
8	42.4	42.0	13842	21	AAZ87297	S. venezuelae macr
9	42.4	42.0	13987	18	AAT80415	Hybrid srmG/tylG O
10	42.4	42.0	36778	21	AAZ87318	S. venezuelae pik
11	42.4	42.0	37948	21	AAZ87285	S. venezuelae pik
12	42.4	42.0	38506	21	AAA75633	Nucleotide sequenc
13	42.4	42.0	38506	21	AAZ56001	Recombinant cosmid
14	42.4	42.0	43280	18	AAT80413	Tylactone synthase
15	42.4	42.0	44377	18	AAT78508	Platenolide syntha
16	42.4	42.0	44377	18	AAT80414	Platenolide syntha
17	41	40.6	990	21	AAC55784	Type I polyketide
18	41	40.6	12249	21	AAC55840	Complete Mitomycin
19	41	40.6	18331	21	AAC55857	Complete nucleotid
20	40.8	40.4	11220	21	AAZ87298	S. venezuelae macr
21	39.8	39.4	4689	21	AAZ87299	S. venezuelae macr
22	39.2	38.8	1681	24	AAS18438	Contig 115 DNA enc
23	39.2	38.8	29879	14	AAQ46806	eryA region of S.
24	38.2	37.8	2237	24	ABL62699	Colon adenocarcino
25	38.2	37.8	3209	21	AAC78080	Human cancer assoc
26	38.2	37.8	8460	18	AAT88206	cDNA for protein (
27	38.2	37.8	8470	21	AAZ95007	Cancer specific ge
28	38.2	37.8	8519	21	AAA37760	Human fatty acid s
29	38.2	37.8	77536	21	AAA14651	Nucleotide sequenc
30	37.6	37.2	15872	21	AAZ87283	S. venezuelae vep
31	37.6	37.2	16767	22	AAF88339	S. spinosa DNA fra
32	37.6	37.2	27541	22	AAD17185	Streptomyces nours
33	37.6	37.2	50000	22	AAF88313	S. spinosa DNA fra
34	37.6	37.2	50000	22	AAF88316	S. spinosa DNA fra
35	37.6	37.2	80161	20	AAZ21501	DNA fragment of Sa
36	37.4	37.0	33529	23	AAS17367	DNA sequence of S.
37	37.2	36.8	109519	22	AAS08693	Micromonospora DNA
38	37	36.6	47981	22	AAF30757	Micromonospora meg
39	36.6	36.2	6564	22	AAF31637	Mycobacterium tube
40	36.6	36.2	9513	22	AAF88337	S. spinosa DNA fra
41	36.4	36.0	5544	22	AAF24891	Pimaricin biosynth
42	36.2	35.8	14775	22	AAF88338	S. spinosa DNA fra
43	36.2	35.8	15872	18	AAT68715	Streptomyces venez
44	36	35.6	2700	22	AAD17193	ERD48 insert DNA i
c 45	36	35.6	14071	23	ABL50560	Micromonospora car

ALIGNMENTS

RESULT 1
AAA29349
ID AAA29349 standard; DNA; 71989 BP.
XX
AC AAA29349;
XX
DT 12-SEP-2000 (first entry)
XX
DE Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /*tag= c
FT /label= epoA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548

```
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                         domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                         domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                         domain, potentially involved in formation of the
FT                         thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                         domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                         module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                         NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                         module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                         module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                         module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                         module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-20;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20490 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 20549

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  20550 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 20590


RESULT 2
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
```

```
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR          AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR          AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR          AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             98.4%;  Score 99.4;  DB 21;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 7.4e-20;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||||||||||||||||||||||||||||||||||||||||
Db   26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 3
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             51.5%;  Score 52;  DB 22;  Length 20394;
  Best Local Similarity   70.0%;  Pred. No. 6e-06;
  Matches   70;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           |||| |  | ||| | ||||  |||||||||  ||||||||||||||| || ||  | |
Db    4469 TCGATGCCCAACGCGCCTTCCGCGAGCAGGGGTTCGACTCCCTGATGGCCGTCGACCTGC 4528

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||    | | |  ||| ||| |||| | ||
Db    4529 GCAACCGGCTCTCCGCCGCGACGGGCCTGCGGCTGCCCGC 4568


RESULT 4
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
```

```
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match             43.6%;  Score 44;  DB 21;  Length 30690;
  Best Local Similarity   65.0%;  Pred. No. 0.0014;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
            ||||||     || |  |||  ||||| |||  ||||||| || | ||| || ||| | |
Db   30161 TCGACGCATCGCGGGCGTTCCGCGAGCTGGGGTTCGACTCGCTCACGGCGGTCGAGCTGC 30220

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| || ||   ||  | |  |||  ||| |||| |||
Db   30221 GCAACCGTCTCACGGCCGCGACGGGCCTGCGGCTGCGGGC 30260


RESULT 5
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             43.6%;  Score 44;  DB 22;  Length 30690;
  Best Local Similarity   65.0%;  Pred. No. 0.0014;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
            ||||||     || |  |||  ||||| |||  ||||||| || | ||| || ||| | |
Db   30161 TCGACGCATCGCGGGCGTTCCGCGAGCTGGGGTTCGACTCGCTCACGGCGGTCGAGCTGC 30220

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| || ||   ||  | |  |||  ||| |||| |||
Db   30221 GCAACCGTCTCACGGCCGCGACGGGCCTGCGGCTGCGGGC 30260


RESULT 6
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
```

```
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match          43.6%;  Score 44;  DB 22;  Length 65140;
  Best Local Similarity   65.0%;  Pred. No. 0.0016;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           |||  | | | |||   |||   |||  |||| ||||||| ||||  || |||||| |||
Db   49425 TCGGTGCGGGGCGACAGTTCCAGGAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCC 49484

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           ||||  ||  ||  |  |  |     ||  ||| |||| ||||
Db   49485 GCAACCGACTCAACGCGGCCACCGGACTGCGGCTGCCGGC 49524


RESULT 7
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
```

```
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             43.6%;  Score 44;  DB 22;  Length 125401;
  Best Local Similarity   65.0%;  Pred. No. 0.0018;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          |||  | | | |||   |||   |||  |||| ||||||| ||||  || ||||||| |||
Db 109686 TCGGTGCGGGGCGACAGTTCCAGGAGTTGGGCTTCGACTCGCTGACCGCGGTGGAGTTCC 109745

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| || ||  | |  |     ||  ||| |||| ||||
Db 109746 GCAACCGACTCAACGCGGCCACCGGACTGCGGCTGCCGGC 109785


RESULT 8
AAZ87297
ID   AAZ87297 standard; DNA; 13842 BP.
XX
AC   AAZ87297;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikAI, SEQ ID NO:30.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..13842
FT                   /*tag=  a
FT                   /product= "PikAI"
FT                   /transl_except= (pos:4156..4158, aa:Ala)
FT                   /transl_except= (pos:13741..14743, aa:Ala)
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77192.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 377-383; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 13842 BP; 1726 A; 5356 C; 4911 G; 1845 T; 4 other;

  Query Match             42.0%;  Score 42.4;  DB 21;  Length 13842;
  Best Local Similarity   64.0%;  Pred. No. 0.0038;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  13286 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13345

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |     ||  |  | ||  ||||
Db  13346 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13385


RESULT 9
AAT80415
ID   AAT80415 standard; DNA; 13987 BP.
XX
AC   AAT80415;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Hybrid srmG/tylG ORF1.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; hybrid gene; ss.
XX
OS   Streptomyces ambofaciens.
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..13987
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
```

```
FT                   /note= "ORF1 encodes hybrid protein shown in AAW22611"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22611.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 22; Pages 178-197; 220pp; English.
XX
CC   This sequence represents a hybrid gene of the invention. This sequence
CC   was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a
CC   EcoRI-SfuI fragment from tylG ORF1. The position of the nucleotides from
CC   each of the two genes is not given in the specification. The srmG gene
CC   (see AAT80414) was isolated from Streptomyces ambofaciens, and encodes
CC   the multi-functional proteins which direct the synthesis of the
CC   polyketide platenolide. Platenolide is the basic building block of the
CC   macrolide antibiotic spiramycin. The tylG gene (see AAT80413) is the
CC   tylactone synthase gene cluster of the invention. The tylG sequence was
CC   isolated from Streptomyces fradiae, and encodes multifunctional proteins
CC   which direct the synthesis of the polyketide tylactone. Tylactone is the
CC   basic building block of the antibiotic tylosin. The hybrid sequence can
CC   be used to transform S. ambofaciens lacking the srmG ORF1 sequence, or S.
CC   fradiae lacking the tylG ORF1 sequence, so that they can produce
CC   polyketides. The DNA sequence can be modified so as to alter the type of
CC   carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 18;  Length 13987;
  Best Local Similarity   64.0%;  Pred. No. 0.0038;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||      ||||  ||||   |||   ||| ||||||||||  ||| ||||||  | |
Db   13551 TCGAGCCCGAACGACCCTTCCGGGAGATCGGCGTCGACTCCCTGGCGGCGGTGGAGCTGC 13610

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||     ||  | | ||  ||| | || || |
Db   13611 GCAACCGGCTCAGCCGTCTGGTCGGCCTGCGGTTGCCGAC 13650


RESULT 10
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 31.
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;

  Query Match             42.0%;  Score 42.4;  DB 21;  Length 36778;
  Best Local Similarity   64.0%;  Pred. No. 0.0044;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
```

```
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  15027 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 15086

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |     ||  |  | ||  ||||
Db  15087 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 15126


RESULT 11
AAZ87285
ID   AAZ87285 standard; DNA; 37948 BP.
XX
AC   AAZ87285;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77180.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 13; Page 299-315; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX

SQ   Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 21;  Length 37948;
  Best Local Similarity   64.0%;  Pred. No. 0.0044;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  16197 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 16256

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |     ||  |  | ||  ||||
Db  16257 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 16296


RESULT 12
AAA75633
ID   AAA75633 standard; DNA; 38506 BP.
XX
AC   AAA75633;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
PR   06-MAY-1998;   98US-0073538.
PR   28-AUG-1998;   98US-0141908.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  Tang L,  McDaniel R;
XX
DR   WPI; 2000-610844/58.
XX
PT   New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT   for converting ketolides to antibiotics and as antibiotics and
PT   intermediates in the synthesis of compounds with pharmaceutical value
PT   -
XX
PS   Disclosure; Columns 15-32; 117pp; English.
XX
CC   The present sequence is used to produce the recombinant DNA compounds
CC   of the invention. The specification describes a recombinant DNA compound
CC   expressing recombinant polyketide synthase genes in host cells for the
CC   production of narbonolide, narbonolide derivatives and polyketides that
CC   are useful as antibiotics and as intermediates in the synthesis of
CC   compounds with pharmaceutical value. The DNA compounds may also encode
CC   a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC   transferase enzymes (useful for conversion of ketolides to antibiotics),
CC   and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC   These compounds are also useful for increasing the antibiotic activity
CC   of a compound relative to the unhydroxylated compound. The recombinant
CC   host cells are useful as genetic systems that allow rapid engineering
CC   of the narbonolide polyketide synthase. These would be valuable for
CC   creating novel ketolide analogs for pharmaceutical applications.
```

```
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 21;  Length 38506;
  Best Local Similarity   64.0%;  Pred. No. 0.0044;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  13169 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13228

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |      ||     | ||  ||||
Db  13229 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13268


RESULT 13
AAZ56001
ID   AAZ56001 standard; DNA; 38506 BP.
XX
AC   AAZ56001;
XX
DT   23-MAR-2000  (first entry)
XX
DE   Recombinant cosmid pKOS023-27 containing S. venezuelae PKS genes.
XX
KW   Narbonolide polyketide synthase; PKS; cosmid pKOS023-27; ketolide;
KW   antibiotic production; narbomycin; picromycin; ds.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             70..13725
FT                   /*tag= a
FT                   /product= PICAI
FT                   /note= "Narbonolide synthase subunit 1"
FT   CDS             13830..25049
FT                   /*tag= b
FT                   /product= PICAII
FT                   /note= "Narbonolide synthase subunit 2"
FT   CDS             25133..29821
FT                   /*tag= c
FT                   /product= PICAIII
FT                   /note= "Narbonolide synthase subunit 3"
FT   CDS             29924..33964
FT                   /*tag= d
FT                   /product= PICAIV
FT                   /note= "Narbonolide synthase subunit 4"
FT   CDS             33961..34806
FT                   /*tag= e
FT                   /product= PicB
FT                   /note= "Contains typeII thioesterase domain"
FT   CDS             34863..36011
FT                   /*tag= f
FT                   /product= PICCII
FT                   /note= "4-keto-6-deoxyglucose isomerase"
FT   CDS             36159..37439
FT                   /*tag= g
FT                   /product= PICCIII
FT                   /note= "Desosaminyl transferase"
FT   CDS             37529..38242
FT                   /*tag= h
FT                   /product= PICCVI
FT                   /note= "3-amino dimethyltransferase"
XX
PN   WO9961599-A2.
XX
PD   02-DEC-1999.
XX
PF   27-MAY-1999;   99WO-US11814.
XX
PR   28-MAY-1998;   98US-0087080.
PR   28-AUG-1998;   98US-0141908.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Ashley G,  Betlach MC,  Betlach M,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-072618/06.
DR   P-PSDB; AAY67201, AAY67202, AAY67203, AAY67204, AAY67205, AAY67207,
DR   AAY67208, AAY67211.
XX
PT   New recombinant DNA encoding a domain of narbonolide polyketide
PT   synthase, for production of ketolide antibiotics -
XX
PS   Example 2; Page 16-27; 98pp; English.
XX
CC   This is the recombinant cosmid pKOS023-27 DNA sequence which contains a
CC   Streptomyces venezuelae DNA insert. The cosmid contains open reading
CC   frames which encode the various modules of the narbonolide polyketide
CC   synthase (PKS). The invention relates to recombinant DNA containing a
CC   coding sequence for a narbonolide PKS. Polyketides are compounds
CC   synthesised from 2-carbon units through a series of condensations and
CC   subsequent modifications. Modular PKSs are responsible for the production
CC   of many antibiotics including picromycin. The narbonolide PKS consists of
CC   a loading module, six extender modules, and two thioester domains. Four
CC   proteins make up the narbonolide PKS (PICAI, PICAII, PICAIII and PICAIV).
CC   PICAI includes the loading module and extender modules 1 and 2, PICAII
CC   includes extender modules 3 and 4, PICAIII includes extender module 5 and
CC   PICAIV includes extender module 6 and a type II thioesterase domain. The
CC   second type II thioesterase domain is found on the PICB protein. The
CC   nucleotide sequences encoding all of these proteins can be isolated in
CC   recombinant form from the recombinant cosmid pKOS023-27. Narbonolide is
CC   desosaminylated in S. venezuelae to yield narbomycin, the desosaminyl
CC   transferase enzyme is required for this  conversion, and the desosamine
CC   biosynthetic genes are also found in cosmid pKOS023-27. The recombinant
CC   DNA of the invention is used to express, in transformed cells,
CC   narbonolide (or its derivatives) or other ketolides (particularly
CC   hybrids), which may then be converted (e.g. by other enzymes
CC   recombinantly expressed in the same hosts) to polyketide antibiotics or
CC   their intermediates. The antibiotics are useful in human or veterinary
CC   medicine.
XX
SQ   Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

  Query Match             42.0%;  Score 42.4;  DB 21;  Length 38506;
  Best Local Similarity   64.0%;  Pred. No. 0.0044;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||||||    ||| ||||||  ||| |  ||| ||||||||||||  || || ||  |||
Db  13169 TCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCC 13228

Qy     62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||| |||||  |    |      ||     | ||  ||||
Db  13229 GCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGC 13268


RESULT 14
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match            42.0%;  Score 42.4;  DB 18;  Length 43280;
  Best Local Similarity  64.0%;  Pred. No. 0.0045;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||||| || ||  |||||   ||   |||| ||||||||||  |  || || ||  | |
Db   19512 TCGACGTCCGCCGGAGCTTCAAGGAAGCGGGCTTCGACTCCCTCACCGCCGTCGAACTGC 19571

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| ||||| |  |  |     ||  ||  |||| | ||
Db   19572 GCAACCGGCTGCGCGCCGCCACCGGGCTGAAGCTGCCCGC 19611


RESULT 15
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match            42.0%;  Score 42.4;  DB 18;  Length 44377;
  Best Local Similarity  64.0%;  Pred. No. 0.0045;
  Matches   64;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           ||||      ||||  ||||   |||   ||| ||||||||||  ||| ||||||  | |
Db   13566 TCGAGCCCGAACGACCCTTCCGGGAGATCGGCGTCGACTCCCTGGCGGCGGTGGAGCTGC 13625

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||     ||  | | ||  ||| | || || |
Db   13626 GCAACCGGCTCAGCCGTCTGGTCGGCCTGCGGTTGCCGAC 13665
```

Search completed: November 5, 2002, 12:37:51
Job time : 273.215 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_20490_20590
Perfect score: 101
Sequence: 1 ctcgacgtgcgacgaggctt.....ctgggtatgccgctgtcggc 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	71989	6	AR172664	AR172664 Sequence
3		99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4		99.4	98.4	68750	6	AR193029	AR193029 Sequence
5		99.4	98.4	68750	6	AR199551	AR199551 Sequence
6		99.4	98.4	68750	6	AR199559	AR199559 Sequence
7		99.4	98.4	68750	6	AR199567	AR199567 Sequence
8		99.4	98.4	68750	6	AR201097	AR201097 Sequence
9		99.4	98.4	68750	6	AR208671	AR208671 Sequence
10		77	76.2	10910	1	AX024383	AX024383 Sequence
11		77	76.2	10910	6	AX024276	AX024276 Sequence
12		52	51.5	20394	1	SNA132222	AJ132222 Streptomy
13		52	51.5	20394	6	AX067996	AX067996 Sequence
14	c	52	51.5	84985	1	SNA278573	AJ278573 Streptomy
15		47.2	46.7	41097	1	AF016585	AF016585 Streptomy
16		44	43.6	30690	6	E38020	E38020 Avermectin
17		44	43.6	64957	1	AB032367	AB032367 Streptomy
18		44	43.6	65140	6	AX211705	AX211705 Sequence
19	c	44	43.6	104326	1	AB070940	AB070940 Streptomy
20		44	43.6	113193	1	AF357202	AF357202 Streptomy
21		44	43.6	123580	1	AF263912	AF263912 Streptomy
22		44	43.6	125401	6	AX211739	AX211739 Sequence
23		42.8	42.4	78210	1	AB070949	AB070949 Streptomy
24		42.4	42.0	37948	1	AF079138	AF079138 Streptomy
25		42.4	42.0	43280	1	SFU78289	U78289 Streptomyce
26	c	42.4	42.0	82746	1	AF453501	AF453501 Actinosyn
27		41.4	41.0	25150	1	AB070945	AB070945 Streptomy
28	c	40.8	40.4	30000	6	AX250263	AX250263 Sequence
29		40.4	40.0	104326	1	AB070940	AB070940 Streptomy
30		39.2	38.8	1681	6	AR173228	AR173228 Sequence
31		39.2	38.8	4015	1	AX024421	AX024421 Sequence
32		39.2	38.8	4015	1	AX024422	AX024422 Sequence
33		39.2	38.8	4015	6	AX024314	AX024314 Sequence
34		39.2	38.8	4015	6	AX024315	AX024315 Sequence
35		39.2	38.8	17101	1	AB017641	AB017641 Micromono
36		39.2	38.8	20235	1	SERERYAB	M63677 S.erythraea
37		39.2	38.8	20235	6	AR049368	AR049368 Sequence
38		39.2	38.8	20235	6	AR095529	AR095529 Sequence
39		39.2	38.8	20444	1	SEERYABS	X62569 S.erythraea
40	c	39.2	38.8	49736	1	AF319998	AF319998 Stigmatel
41		38.8	38.4	36181	1	U00023	U00023 Mycobacteri
42	c	38.8	38.4	308950	1	MLEPRTN9	AL583925 Mycobacte
43		38.6	38.2	59816	1	AF333038	AF333038 Streptomy
44		38.2	37.8	2237	6	AX330527	AX330527 Sequence
45		38.2	37.8	2237	9	S80437	S80437 fatty acid

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L , Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:55"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGIVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNIKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDPERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
     gene            6260. .10492
                     /gene="epoB"
     CDS             6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
     gene            10639. .16137
                     /gene="epoC"
     CDS             10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDFALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20490 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 20549

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
             |||||||||||||||||||||||||||||||||||||||||
Db   20550 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 20590


RESULT 2
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others

ORIGIN

  Query Match          100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20490 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 20549

Qy      61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
             |||||||||||||||||||||||||||||||||||||||||
Db   20550 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 20590


RESULT 3
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
```

```
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETIGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALIRFWRSVYVATESA"
CDS                  3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown'
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHEALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVFAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAFVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
CDS                  complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
CDS                  6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYAWFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene                 7610. .11875
                     /gene="epoA"
CDS                  7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature         7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
misc_feature         9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
misc_feature         10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
misc_feature         11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
gene                 11872. .16104
                     /gene="epoP"
CDS                  11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene                 16251. .21749
                     /gene="epoB"
CDS                  16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

Query Match          98.4%;  Score 99.4;  DB 1;  Length 68750;
```

```
  Best Local Similarity   99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 4
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 5
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 6
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 7
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202
```

```
RESULT 8
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity  99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 9
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity  99.0%;  Pred. No. 2.9e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db  26102 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCCGTGGAGATC 26161

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          |||||||||||||||||||||||||||||||||||||||||
Db  26162 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 26202


RESULT 10
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           76.2%;  Score 77;  DB 1;  Length 10910;
  Best Local Similarity  85.1%;  Pred. No. 1.4e-09;
  Matches   86;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||  | ||||||||||||||||||| |||||||||||||||||| ||||||||
Db   2037 CTCGACGTCGGCCGAGGCTTCGCCGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATC 2096

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          || || || |||||| | ||||||||     | |||||||||
Db   2097 CGTAACCGCCTTCAGCGCGAGCTGGGCGAACGGCTGTCGGC 2137


RESULT 11
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match           76.2%;  Score 77;  DB 6;  Length 10910;
  Best Local Similarity  85.1%;  Pred. No. 1.4e-09;
  Matches   86;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 CTCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATC 60
          ||||||||  | ||||||||||||||||||| |||||||||||||||||| ||||||||
Db   2037 CTCGACGTCGGCCGAGGCTTCGCCGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATC 2096

Qy     61 CGCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
          || || || |||||| | ||||||||     | |||||||||
Db   2097 CGTAACCGCCTTCAGCGCGAGCTGGGCGAACGGCTGTCGGC 2137


RESULT 12
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
```

KEYWORDS pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 20394)
AUTHORS Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
TITLE The biosynthetic gene cluster for the 26-membered ring polyene macrolide pimaricin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes
JOURNAL J. Biol. Chem. 274 (15), 10133-10139 (1999)
MEDLINE 99214571
PUBMED 10187796
REFERENCE 2 (bases 1 to 20394)
AUTHORS Aparicio,J.F.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av. Real, 1-Leon, 24006, SPAIN
FEATURES Location/Qualifiers
source 1. .20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
gene 1. .20394
/gene="pimS1"
CDS 1. .20394
/gene="pimS1"
/function="pimaricin biosynthesis"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="CAB41041.1"
/db_xref="GI:4678703"
/translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDLVFAEEGSAEAALLD
QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG

```
                    LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                    DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                    VGPEALLGELERMEKSFSGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                    DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT      2611 a    7486 c    7397 g    2900 t
ORIGIN

  Query Match             51.5%;  Score 52;  DB 1;  Length 20394;
  Best Local Similarity   70.0%;  Pred. No. 0.0023;
  Matches    70;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          |||| |  | ||| | ||||  |||||||||  ||||||||||||||||| || ||  | |
Db    4469 TCGATGCCCAACGCGCCTTCCGCGAGCAGGGGTTCGACTCCCTGATGGCCGTCGACCTGC 4528

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||    |  | |  |||  | | |||| | ||
Db    4529 GCAACCGGCTCTCCGCCGCGACGGGCCTGCGGCTGCCCGC 4568


RESULT 13
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24542.1"
                     /db_xref="GI:12329307"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAIPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTECVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGREPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDREEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVFPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVIDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRFAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLFDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
```

```
                     HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                     LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                     DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                     VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                     DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a   7486 c   7397 g   2900 t
ORIGIN

  Query Match           51.5%;  Score 52;  DB 6;  Length 20394;
  Best Local Similarity   70.0%;  Pred. No. 0.0023;
  Matches   70;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           |||| |  | ||| | ||||  |||||||||  |||||||||||||||| || ||  | |
Db    4469 TCGATGCCCAACGCGCCTTCCGCGAGCAGGGGTTCGACTCCCTGATGGCCGTCGACCTGC 4528

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||    |  | |  |||  ||| |||| | ||
Db    4529 GCAACCGGCTCTCCGCCGCGACGGGCCTGCGGCTGCCCGC 4568


RESULT 14
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
     gene            2190. .3566
                     /gene="pimK"
     CDS             2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
     gene            complement(3662. .9736)
                     /gene="pimS4"
     CDS             complement(3662. .9736)
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                     REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                     DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                     GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                     AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                     LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                     MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
```

```
                SMLRGLVKGARRTAVNITGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                QNVDARREFYELGFDSITSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                ELEDLPLPATLAEGPGIPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772. .15198)
                /gene="pimS3"
CDS             complement(9772. .15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVSMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHFWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGIAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNITGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLITSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
```

```
                DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
                PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT
```

```
  Query Match             51.5%;  Score 52;  DB 1;  Length 84985;
  Best Local Similarity   70.0%;  Pred. No. 0.002;
  Matches   70;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
           |||| |  | ||| | |||| |||||||| ||||||||||||||| || ||  | |
Db   76684 TCGATGCCCAACGCGCCTTCCGCGAGCAGGGGTTCGACTCCCTGATGGCCGTCGACCTGC 76625

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||    |  | |  |||  ||| |||| | ||
Db   76624 GCAACCGGCTCTCCGCCGCGACGGGCCTGCGGCTGCCCGC 76585


RESULT 15
AF016585
LOCUS       AF016585               41097 bp    DNA     linear   BCT 07-DEC-1997
DEFINITION  Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi)
            gene, partial cds; polyketide synthase modules 1 through 7 (nidA)
            genes, complete cds; and N-methyltransferase homolog gene, partial
            cds.
ACCESSION   AF016585
VERSION     AF016585.1  GI:2558836
KEYWORDS    .
SOURCE      Streptomyces caelestis.
  ORGANISM  Streptomyces caelestis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 41097)
  AUTHORS   Kakavas,S.J., Katz,L. and Stassi,D.
  TITLE     Identification and characterization of the niddamycin polyketide
            synthase genes from Streptomyces caelestis
  JOURNAL   J. Bacteriol. 179 (23), 7515-7522 (1997)
  MEDLINE   98053867
   PUBMED   9393718
REFERENCE   2  (bases 1 to 41097)
  AUTHORS   Kakavas,S. and Stassi,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park
            Road, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .41097
                     /organism="Streptomyces caelestis"
                     /strain="NRRL-2821"
                     /db_xref="taxon:36816"
     gene            complement(<1. .396)
                     /gene="nidi"
     CDS             complement(<1. .396)
                     /gene="nidi"
                     /note="ORF6; similar to tyli"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P-450 hydroxylase homolog"
                     /protein_id="AAC46023.1"
                     /db_xref="GI:2558837"
```

```
                /translation="MVDSVTGPMELSKDANAKELLDWFSHNRTHHPVFWDEGRQAWQV
                FRYDDYLTVSNHPEFFSSDFTEVAPTPPELEMILGPGTIGALDPPAHGPMRKLVSQAF
                TPRRMAGQEQRIRVIAEELLDRVRGQKTIA"
gene            627. .13649
                /gene="nidA1"
CDS             627. .13649
                /gene="nidA1"
                /note="orf1"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase modules 1 and 2"
                /protein_id="AAC46024.1"
                /db_xref="GI:2558838"
                /translation="MAGHGDATAQKAQDAEKSEDGSDAIAVIGMSCRFPGAPGTAEFW
                QLLSSGADAVVTAADGRRRGTIDAPADFDAAFFGMSPREAAATDPQQRLVLELGWEAL
                EDAGIVPESLRGEAASVFVGAMNDDYATLLHRAGAPTDTYTATGLQHSMIANRLSYFL
                GLRGPSLVVDTGQSSSLVAVALAVESLRGGTSGIALAGGVNLVLAEEGSAAMERVGAL
                SPDGRCHTFDARANGYVRGEGGAIVVLKPLADALADGDRVYCVVRGVATGNDGGGPGL
                TVPDRAGQEAVLRAACDQAGVRPADVRFVELHGTGTPAGDPVEAEALGAVYGTGRPAN
                EPLLVGSVKTNIGHLEGAAGIAGFVKAALCLHERALPASLNFETPNPAIPLERLRLKV
                QTAHAALQPGTGGGPLLAGVSAFGMGGTNCHVVLEETPGGRQPAETGQADACLFSASP
                MLLLSARSEQALRAQAARLREHLEDSGADPLDIAYSLATTRTRFEHRAAVPCGDPDRL
                SSALAALAAGQTPRGVRIGSTDADGRLALLFTGQGAQHPGMGQELYTTDPHFAAALDE
                VCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARGTQAHLVL
                GHSVGEITAAHIAGVLDLPDAARLITARAHVMGQLPHGGAMLSVQAAEHDLDQLAHTH
                GVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKVSHAFHSALMDPMLGAFR
                DTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQTAHHQGTT
                TYLEIGPHPTLTTLLHHTLDNPTTIPTLHRERPEPETLTQAIAAVGVRTDGIDWAVLC
                GASRPRRVELPTYAFQRRTHWAPGLTPNHAPADRPAAEPQRAMAVGPVSREALVRLVG
                ETTASVLGLDGPDEVALDRPFTSQGLDSMTAVELAGLLGTAAGVALDPTLVYELPTPR
                AVADHLAKTLLGESAADADQEVNGRTGEAEAKAGDPIAVIGIGCRFPGGVATPDDLWE
                LVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQFDAEFFGISPR
                EATAMDPQQRLLLETSWEALERAGVVPESLRGGRTGVFVGTTAPEYGPRLHEGTDGYE
                GFLLTGTTASVASGRIAYALGTRGPALTVDTACSSSLVALHLAVQSLRRGECDLALAG
                GTTVMSGPGMFVEFSRQRGLAPDGRCKAFSADADGTAWAEGVGMLLVERLSDAERLGH
                RVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDAVEAHGTGTAL
                GDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGIIKMVQALRHDTLPRT
                LHADHPSSKVDWDAGPLQLLTDARPWPADPDRPRRAGISAFGVSGTNAHVVLEEPPAP
                EEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIAHALATTRTHF
                KHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHRGMGQELYATDPH
                FAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARG
                TQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHGGAMLSVQAAEHDL
                DQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQQLHQQGIRHTWLKVSHAFHSALMD
                PMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQT
                AHDQRTTTYLEIGAHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTALATLHTTGHT
                TTLHTTSPQTHHLDLPTYPFQRDRYWMEPVRVAQVSGQPGADRLRYRVVWEAAPEDHT
                VPSAERWLLLGGSDPADAGILVGVEEQLVGHGAAVQRLDIPAPAGRRLLPSCSELPVR
                GPSHTGEFCGSRLRSPRVAEAVALLQALGDAGADAPLWIATRGAVAAQAGEAPSVGGA
                QLWGLGQVAGLELADRWGGLVDLPADPAPAALRGLARVLVSNAPDNQVAIRASGVFVR
                RVVPAPGRPVRNDWAPSGTVLITGGTGALGSQVARRLALAGAPHLLLAGRRGNSGAAE
                LVDELTALGAEVTVAACDAADRDALVSLLATIPEHRPLTAVLHAAGVLDDGVLDGLTP
                ERIDAVLRAKATAARHLDELTADLDLDAFVLFSSIVGVWGNGGQATYAAANAALDALA
                HGRRARGQRATSIAWGPWAGSGMAAGDGAKSLARDGIDVLEPEEALDLLEQAVRSDET
                TLLVADVDWETFLSSSASRRFQLPLFGRIPAARAVAVSGADGSPDDVPVWLSPQLSAA
                ERQLRLLDLVRTEAASVLRHGSATAVDAETAFRAGGFDSLTLVELRNRLASATGLRLS
                NTLLFDHPTPAAVAEYLGEQLLGTEPSAAASHGDAGARTVVASDEPVAVVGMACRYPG
                GVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQF
                DAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGASQQEYGT
                QSREAADKYGGHLLTGTLASVMSGRVAYTLGLQGPALTVDTACSSSLVALHLAVQSLR
                RGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVGVLLVE
                RLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDA
                VEAHGTGTPLGDPIEAEALLATYGHPERQTPVWLGSLKSNIGHSQAAAGIAGIIKMIQ
                AMHHGTLPRTLHVDAPTSQVDWEAGSLQLLAEARPWPADPDRPRRAGISAFGVSGTNA
                HVVLEEPPAPEEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIA
                HALATTCTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHPGM
                GQELYTTDPHFAAALDEICEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQT
                ALYRTLTAHGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHDGAM
                LSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKV
                SHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARH
                TVRFADAVQTAHDQRTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTA
                LATLHTTGHTTTPHPSHIPAQRVSLPAYPFQRRAYWMPNSAARIGRSDAEAATRLGLA
                RQDHPFLTGATSIVGSGATLLTGRVSLAAHPWLADHTVAGAVLFPGTAFADLLLHVAE
                ETGSGTVEELTLHVPLVLPAQGAVQLQVLVEAPDERGRRAATVAARPEEETGAEWTTH
                AAGVLAGAEPAAPAVDWADWAAGAWPPPGAEAVDVDELYAEFAAAGYDYGPAFTGLTG
                AWRRGEEVFADVRLPEAVAGSGEEERRFGVHPALFDAALHPWRVGGPWQESGPDADGG
                QSTLLPFSWQGLTLYGTGAEQVRVRLAPSDGGAFSVRAADPAGHPVLALDALVLRRVP
                SSTPGALPAAPMYHIDWQPVDSTSLVDTRRYAVVGADADETAEVLRAFLPDAHTTAHA
                DLAALRAAPDAVAPTPGPALVVALIPGSPPPAPEATSAPDPVRATLAHGLSLVQALIE
                DERLTTGTTLAVVTRGAVEATQTDVPDLAGAALWGLVRSAQSEYPGRFRLIDVDDSPE
                SRAALATALASDEPQLALRAGVMLAPALTPAEAAGSPAPADAAPTAHGMDGTVLLTGC
                TGALGRRVAVHLARRHGVRRMLLVSRSGPNAPESAALERELTELGVHTSFVSCDLADP
                ADVRKAVAGVPSDHPLTGVVHTAGVLDDGALARLTPERIDTVLRAKADAVRHLHEATL
                GHPLRAFVLFSAAAGLLGRPGQASYAAANAVLDAFARLRRAEGLPAVSLAWGLWDERE
                GMAGGLDATALRRLRRDGILPMPADQALELLDQALTTYRDGPALLVPLLLDSTALRRT
                AGENGPASVPPLLRSLVPSHPRRHRTTTATSVRDESGPGPSARLAALPAKERMTVLLD
                TVSEQVAGVLGHASAAEIDPERPFHEIGFDSLATLELRNRLGRLVGLRLPSTLAFDHP
                TPRKVAEWINAELPAANPALGTSAEDPVLTGIDQMARAIALMDAEDGRRADVRQRLVG
                LLTALDRPASGADSTRAAATTVADRLDEATDDEIFAFLDEQLEEQPDEPLGAPLDERL
                DEQP"
gene            13693. .19212
                /gene="nidA2"
CDS             13693. .19212
                /gene="nidA2"
                /note="orf2"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase module 3"
                /protein_id="AAC46025.1"
                /db_xref="GI:2558839"
                /translation="MTAENDKIRSYLKRATAELHRTKERLAELESARHEPVAVVGMAC
                RYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHD
                AAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPNTLRTTRTGVFTGVMYD
                DYGSRCFPSLPEVEGYVVNGSAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLAV
                QSLRRGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKSFADTADGTAWAEGVGV
                LLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSAD
                DIDAVEAHGTGTPLGDPIEAGALLATYGHPERQTPVWLGSLKSNIGHTQAAAGIAGII
                KMIQAMHHGTLPRTLHVDHPSSKVDWKAGSLQLLAQARPWPADPDRPRRAAVSAFGVS
                GTNAHAIIEEPSTLIEAPVPDETPVPAHTPPVIAWPLSAHTPTALRAQAARLDSWLHH
                TQADPQDIAHALATTRTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFT
                GQGAQHPGMGQELYTTDPHFAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYT
                QPALFALQTALYRTLTARGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLM
                GQLPHGGAMLSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQ
                NVRHTWLKVSHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTP
                DYWIDHARHTVRFADAVQTAHHQGTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREH
                PEPETLTTALATLHTTGHTTTLHTTSPQSHHLDLPTYPFQRDRYWMAVPPRAAVGDLA
                AAGVTAVGHPLLAAVTEVPETDTTVWTGMLSARALPWLADHLVWDRGVAPGTAVLEIV
                LEIGERIGLPCIGELVLEAPLSWGADAPAVVRTVVGALSPEGTREVSLYSRPEDEAVW
                TRHASGRLARAGSEEMGSGVFAELSGVWPPPGAQEVDISGQYELFAAAGVGYGGAFRG
                LRAVWRREDEIFAEVRLPDDHATDAAGYGLHPALLDAALHPIALLDPLGDGGHGLLPF
                VWTDVLRPGNGGHTLRVRVTRAQDPADGGQDDRPGGAVSLAAADRDGVPVFVARSLVL
                RRVSATQLPAAPFVPLHRLDWVPIPGFTESTGAGGSDPSVLDLTELTADVHLTVREVL
                EAVQAHLREPVGDGGRLVVLTRGAAGPEGGADAPAGAAAWGLVRSAQSEHPGRFVLID
                VDGTEESSRALPAVMAAVGDVDAPQVAVRQGRPSVPRLVPLTADRTEAEPGGPGSAGL
                LGPDGTVLITGGTGALAAVTARHLVARHGARHLLLVSRRGTDAPQAAELASELGALGA
                DVEFRACDVTDRAAVQRLLAGIPAEHPLSCVVHTAGVLDDGVISAQTPERVAAVLGPK
                ADGALHLDELTRDRERSVPLVLFSSVSGVLGSTGHAGYAAANAFLDALAARRRADGHP
                ALSLGWGWWTGTGMAAGLDRADTARIRRSGIAPLDAETAMDLFDRALARPEPALLPVR
                LDHAALREAARAGTLPEVLRDLAGVPSRAHPAPGSEGTATSTGTGSGLAERLARRTVA
                ERLPLLLELVRTEVAAVLGHGDPGAVGAERSFKDAGFDSLTAVDLRNRLNARTGLRLP
                ATLVFDHPTPLALAELLLAGLDVEPAADGPEDLAAAFDRLARGLAALDDRAARVGAAE
                RLRSLLTALTSADPATGTGAQAGPEADGALTEVVTDRLRSASDDELFDLYDSEFR"
gene            19269. .30233
                /gene="nidA3"
CDS             19269. .30233
                /gene="nidA3"
                /note="orf3"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase modules 4 and 5"
                /protein_id="AAC46026.1"
                /db_xref="GI:2558840"
                /translation="MSATNEEKLREYLRRAMGDLHSARERLAELESARHEPVAVVGMA
                CRYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLH
                DAAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGVMH
                HDYGSHQVGSAADASGQLGLGTAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLA
                VQSLRRGECDLALAGGTTVMATPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVG
```

MLLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALADAHLTP
DDIDAVEAHGTGTPLGDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGI
IKMVQALRHDTLPRTLHADHPSSKVDWEAGPLQLLAQARPWPADPDRPRRAGISAFGV
SGTNAHAIIEEPPTLTPAPADTHTPPVIAWPLSAHTPTALRAQAARLRARLHRTSDAL
SLADSAGIGHGLAVGRAALPHRAVLLGDGAAPLDALAALASGEVSPDVVTGSAADVRR
VAFVFPGQGAQWAGMGAELLDSSPVFAAELARCEAALEPFVDWSLTDVLRGAPGAPGL
DRVDVVQPVTFAVVVALAAMWRWLGVEPAAVVGHSQGEIAAAHVAGVLSLEDAARVVA
LRSQLIARELAGRGSMASVALAAADVESRLAGAEAGGGVRDVEIAAVNGPETTVVCGA
PGAVDSLLGVLQGEGVFVRRIDVDYASHSRHVEGIRDELAAVLAGLRPRAGRVPFYST
VEAEPLDGTALDAGYWYRNLRQRVRFESALRAMLADGVDAFVECSPHPVLTVPVRQTL
EDAGAGAVAVGSLRRDEGGLRRFLTSAAEAQVAGVPVDWAALCPRAGWVDLPTYAFQR
ERYWVAPAEPGPAAGAGSAAATGPAAATAARSGSGQEDAASAYGDRLAYHVAWQGLLP
GADAWRPGPRLLIVPAGEREAEVADSVEQAIASFGGTVRRVTADPLSTGRDGLRALLA
PVVAGEEPLTGAVSLLGLCTDGHPEHPGVPAGVTGTLALVQALADLGVAVPLWTVTRG
TVLVVGDSTTVPARLLRSLLDDGAERVVLAGADAGTATDAGTAELAPDAAVVRIPCDV

Query Match 46.7%; Score 47.2; DB 1; Length 41097;
Best Local Similarity 67.0%; Pred. No. 0.034;
Matches 67; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy       2 TCGACGTGCGACGAGGCTTCGCCGAGCAGGGCCTCGACTCCCTGATGGCTGTGGAGATCC 61
          ||| |  |   || | ||||  ||||  |||| ||||||||||||  || |||||  | |
Db   29776 TCGCCTCGGACCGGGCCTTCCGCGAGGCGGGCTTCGACTCCCTGACCGCGGTGGAACTGC 29835

Qy      62 GCAAACGGCTTCAGGGTGAGCTGGGTATGCCGCTGTCGGC 101
           |||| |||||    |  | |  |||| ||| |||| | ||
Db   29836 GCAACCGGCTGGCCGCGGCGACGGGTCTGCGGCTGCCCGC 29875
```

Search completed: November 5, 2002, 12:14:17
Job time : 611.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050
Perfect score: 101
Sequence: 1 gcatgggtgctgcacaccct.....tcgatcggccagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	35.6	35.2	1600	12	BG025615	BG025615 602274546
c	2	31.4	31.1	924	12	BG481933	BG481933 602526850
	3	30.8	30.5	512	10	BE481325	BE481325 166727 BA
	4	30.6	30.3	338	10	BB868343	BB868343 BB868343
c	5	30.6	30.3	921	14	BM817240	BM817240 HC108H04_
	6	30.6	30.3	1101	17	CNS016HE	AL106748 Drosophil
c	7	30.4	30.1	517	13	BI338379	BI338379 362137 MA
c	8	30.2	29.9	379	9	AA917322	AA917322 on45b01.s
	9	30.2	29.9	440	14	C19416	C19416 C19416 Rice
	10	30.2	29.9	571	14	BQ172791	BQ172791 1008b05 r
c	11	30.2	29.9	1146	14	BM920552	BM920552 AGENCOURT
c	12	30	29.7	455	10	BE495030	BE495030 WHE1267_F
c	13	30	29.7	858	13	BI152354	BI152354 602917523
c	14	30	29.7	988	13	BM452552	BM452552 AGENCOURT
c	15	29.8	29.5	371	9	AA081455	AA081455 zn04f10.s
	16	29.8	29.5	433	10	BB696825	BB696825 BB696825
	17	29.8	29.5	444	14	W59094	W59094 md74d02.r1
	18	29.8	29.5	466	9	AI853799	AI853799 UI-M-BH0-
c	19	29.8	29.5	485	10	AW049268	AW049268 UI-M-BH1-
	20	29.8	29.5	493	10	BE647596	BE647596 UI-M-BH1-
c	21	29.8	29.5	499	13	BM220524	BM220524 C0942A03-
	22	29.8	29.5	502	14	W91200	W91200 mf72g08.r1
	23	29.8	29.5	696	10	BB522288	BB522288 BB522288
	24	29.6	29.3	289	12	BF146221	BF146221 WHE1839_F
c	25	29.6	29.3	375	10	BE358851	BE358851 DG1_32_D1
c	26	29.6	29.3	509	10	BE358824	BE358824 DG1_32_D1
	27	29.6	29.3	983	17	CNS03EFY	AL240343 Tetraodon
c	28	29.6	29.3	1047	17	CNS03CQ6	AL238119 Tetraodon
c	29	29.4	29.1	175	9	AU067863	AU067863 AU067863
	30	29.4	29.1	381	14	BM953382	BM953382 952061F02
c	31	29.4	29.1	541	12	BE725405	BE725405 894082H12
	32	29.4	29.1	577	12	BE725525	BE725525 894083H12
c	33	29.4	29.1	584	9	AL506602	AL506602 AL506602
	34	29.4	29.1	590	14	BM952881	BM952881 952059H09
	35	29.4	29.1	665	12	BG860935	BG860935 1024073F1
c	36	29.4	29.1	749	12	BG860934	BG860934 1024073F1
c	37	29.2	28.9	403	9	AA856344	AA856344 vw99g01.r
	38	29	28.7	316	10	BB841966	BB841966 BB841966
	39	29	28.7	333	10	BB844469	BB844469 BB844469
	40	29	28.7	360	10	BB844107	BB844107 BB844107
	41	29	28.7	362	10	BB844104	BB844104 BB844104
	42	29	28.7	384	10	BB844184	BB844184 BB844184
	43	29	28.7	456	10	BB852664	BB852664 BB852664
	44	29	28.7	472	10	AV391765	AV391765 AV391765
	45	29	28.7	482	10	BB853833	BB853833 BB853833

ALIGNMENTS

RESULT 1
BG025615/c
LOCUS BG025615 1600 bp mRNA linear EST 24-JAN-2001
DEFINITION 602274546F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362754 5', mRNA sequence.
ACCESSION BG025615
VERSION BG025615.1 GI:12412391
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1600)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis Staudt, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10007 row: k column: 11
High quality sequence stop: 657.
FEATURES Location/Qualifiers
source 1. .1600

```
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4362754"
                     /clone_lib="NIH_MGC_85"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.867 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      334 a    557 c    558 g    150 t      1 others
ORIGIN

  Query Match           35.2%;  Score 35.6;  DB 12;  Length 1600;
  Best Local Similarity   62.2%;  Pred. No. 6;
  Matches   56;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       7 GTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCG 66
           | ||||  ||||  | |||| ||| |||||||||  |||  |   |     | ||||||
Db     606 GAGCTGGCCACCTGTTCCCGTGAGGAGCCGCTGGTACTCGACAGCCCCGCGTTCTCGGCC 547

Qy      67 TCGGGCGTCTTCGGCTCGATCGGCCAGGGC 96
           | | ||||||||  || |  |  |  |  |
Db     546 TTGAGCGTCTTCACCTTGTCCTCCAGGCGC 517


RESULT 2
BG481933/c
LOCUS       BG481933                924 bp    mRNA    linear   EST 21-MAR-2001
DEFINITION  602526850F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4650204 5',
            mRNA sequence.
ACCESSION   BG481933
VERSION     BG481933.1  GI:13414212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1431  row: d  column: 13
            High quality sequence stop: 835.
FEATURES             Location/Qualifiers
     source          1. .924
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4650204"
                     /clone_lib="NIH_MGC_21"
                     /tissue_type="choriocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                     for average insert size 1.8kb.  Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      186 a    275 c    305 g    157 t      1 others
ORIGIN

  Query Match           31.1%;  Score 31.4;  DB 12;  Length 924;
  Best Local Similarity   61.7%;  Pred. No. 78;
  Matches   50;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           |||| ||| || ||| ||   |   |   |  ||||  ||||||| ||||||| | | |
Db     749 CTGCCCACACTGACCTGCCCTCCCACTTGGCCCCTCAGCGTACTGATTTCCTCCTCATGG 690

Qy      70 GGCGTCTTCGGCTCGATCGGC 90
           | ||||| | | |  | ||
Db     689 TTCTTCTTCAGGTAGGCCAGC 669


RESULT 3
BE481325
LOCUS       BE481325                512 bp    mRNA    linear   EST 28-AUG-2000
DEFINITION  166727 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE481325
VERSION     BE481325.1  GI:9600858
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
  TITLE     Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sonstegard TS
            USDA, ARS, Beltsville Agricultural Research Center
            Bdlg. 200 Rm 2A, Beltsville, MD 20705, USA
            Tel: 301 504 8416
            Fax: 301 504 8414
            Email: tads@anri.barc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 9  row: M  column: 23
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="BARC 5BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."
BASE COUNT      158 a    127 c    130 g     97 t
ORIGIN

  Query Match           30.5%;  Score 30.8;  DB 10;  Length 512;
  Best Local Similarity   58.9%;  Pred. No. 98;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       9 GCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTC 68
           |||| | | ||    || | ||||    |||| |  ||||    |||| | ||  |  ||
Db     223 GCTGAATATCCAGGGCCTCAAGCATGACCTGGGCGACTTCCCCGTGTTCTTCTTCGGCTC 282

Qy      69 GGGCGTCTTCGGCTCGATCGGCCAGGGCAG 98
             || || |  || | ||  |||||||||
Db     283 TCACGACTACTACTGGGTCCACCAGGGCAG 312


RESULT 4
```

```
BB868343
LOCUS       BB868343                338 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB868343 RIKEN full-length enriched, 8 days neonate male testis Mus
            musculus cDNA clone G630004B09 5', mRNA sequence.
ACCESSION   BB868343
VERSION     BB868343.1  GI:17114553
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 338)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .338
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="G630004B09"
                     /clone_lib="RIKEN full-length enriched, 8 days neonate
                     male testis"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="8 days neonate"
                     /note="pooled tissues ; (tissue_type=cerebellum,
                     dev_stage=16 days neonate, sex=mixed),
                     (tissue_type=cerebellum, dev_stage=0 day neonate,
                     sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
                     sex=male), (tissue_type=whole body, dev_stage=9 days
                     embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
                     embryo, sex=mixed)"
BASE COUNT       68 a     96 c    134 g     40 t
ORIGIN
```

```
  Query Match           30.3%;  Score 30.6;  DB 10;  Length 338;
  Best Local Similarity   56.4%;  Pred. No. 1e+02;
  Matches   57;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||  ||| |  | ||||| |   |  |     |  |   || ||| ||   ||| ||  |
Db    161 GCGGGGGAGGCGGACACCATGGGCAACCGCGGGATGGAAGAGCTCATCCCGCTGGTTAAC 220

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
            |  |  || || |||| |||| ||||| ||| || || |
Db    221 AAGCTGCAGGACGCCTTCAGCTCCATCGGTCAGAGCTGCCA 261


RESULT 5
BM817240/c
LOCUS       BM817240                921 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  HC108H04_SK.ab1 HC Hordeum vulgare cDNA clone HC108H04_SK.ab1, mRNA
            sequence.
ACCESSION   BM817240
VERSION     BM817240.1  GI:19153254
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Ozturk,N.Z., Michalowski,C.B., Brazille,S., Borchert,C., Palacio,C.
            , Normand,C., Murphy,C., Kelley,R., Sant,S.A., McLaughlin,H.,
            Fredricksen,M.A. and Bohnert,H.J.
  TITLE     Monitoring large-scale changes in transcript abundance in drought-
            and salt-stressed barley
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Mark A. Fredricksen
            Plant Biology
            University of Illinois
            1201 W Gregory Dr, Urbana, IL 61801, USA
            Tel: 2172655473
            Email: bohnertlab@life.uiuc.edu.
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="Hordeum vulgare"
                     /strain="cv tokak"
                     /db_xref="taxon:4513"
                     /clone="HC108H04_SK.ab1"
                     /clone_lib="HC"
                     /tissue_type="Root"
                     /dev_stage="3 week old"
                     /note="6 and 10 hour drought stress by placing plants on
                     moist paper (75% rel. humidity) in light"
BASE COUNT      270 a    257 c    226 g    168 t
ORIGIN

  Query Match           30.3%;  Score 30.6;  DB 14;  Length 921;
  Best Local Similarity   62.3%;  Pred. No. 1.3e+02;
  Matches   48;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy     10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
          || |||||| ||  | || ||| |||| ||   |||  || | |||   || |||||
Db    349 CTTCACACCATTTTCGGCCAGCTGCCGTTGTTGCTCGGCGAAATGTTGGGCTTGGCGTGC 290

Qy     70 GGCGTCTTCGGCTCGAT 86
          ||  ||   || | | |
Db    289 GGTCTCGGTGGTTTGTT 273


RESULT 6
CNS016HE
LOCUS       CNS016HE                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16M12 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106748
```

```
VERSION     AL106748.1  GI:5623614
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN16M12"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      239 a    215 c    191 g    239 t    217 others
ORIGIN

  Query Match             30.3%;  Score 30.6;  DB 17;  Length 1101;
  Best Local Similarity   44.7%;  Pred. No. 1.4e+02;
  Matches   38;  Conservative   18;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       9 GCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTC 68
           ||||| | |:| : |||| : :: :  :  ||   | ||::| :|: :||||||:|| |:
Db     860 GCTGCCCCCSCCYSCCCGGSSSSGSGGSGGGGGGKTKTTYSTTSTSCKTTCCTCSGCCTY 919

Qy      69 GGGCGTCTTCGGCTCGATCGGCCAG 93
           :||:    | :|:|    |||| |
Db     920 CBGCSGGGCCYKCBCCTCTGGCCTG 944


RESULT 7
BI338379/c
LOCUS       BI338379                517 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  362137 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI338379
VERSION     BI338379.1  GI:15031662
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 130  row: K  column: 20
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .517
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT      107 a    126 c    186 g     98 t
ORIGIN

  Query Match             30.1%;  Score 30.4;  DB 13;  Length 517;
  Best Local Similarity   57.3%;  Pred. No. 1.3e+02;
  Matches   55;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy       2 CATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCT 61
           | ||  ||  ||  |      |||  |||  ||| || |||| || |  ||||| |||
Db     207 CTTGTCTGAGGCCTAGGAGCCCCCAGGAGGTGCCCCTCGACCCCTGCTCACTGTTCTCCA 148

Qy      62 CGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCA 97
             |   | | |  ||||| || |  | | |  |||||
Db     147 GGCGCTGGAGGCTCTTCCGCCCCTTGGCCTTGGGCA 112


RESULT 8
AA917322/c
LOCUS       AA917322                379 bp    mRNA    linear   EST 23-JUN-1998
DEFINITION  on45b01.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1559593 3'
            similar to gb:M16538 GUANINE NUCLEOTIDE-BINDING PROTEIN G(I)/G(S
            )/G(T) BETA SUBUNIT (HUMAN);, mRNA sequence.
ACCESSION   AA917322
VERSION     AA917322.1  GI:3057212
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html

            Trace considered overall poor quality
            Insert Length: 552   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 1.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1559593"
                     /clone_lib="NCI_CGAP_Co8"
                     /tissue_type="adenocarcinoma"
                     /lab_host="DH10B"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; 1st strand cDNA was prepared from
```

colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 98 a 100 c 110 g 71 t
ORIGIN

Query Match 29.9%; Score 30.2; DB 9; Length 379;
Best Local Similarity 65.7%; Pred. No. 1.3e+02;
Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

```
Qy      28 GAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATC 87
           ||| ||| |  || |||| || | ||| || ||| ||  | |||| || ||  |||
Db     169 GAGGAGCAGGAGGCCCTCATCCTTCTGCTTCCCTGGGGTTGGGGCCTCATCCCCTCTGGA 110

Qy      88 GGCCAGG 94
           |||| ||
Db     109 GGCCCGG 103
```

RESULT 9
C19416
LOCUS C19416 440 bp mRNA linear EST 03-APR-2002
DEFINITION C19416 Rice panicle at ripening stage Oryza sativa (japonica cultivar-group) cDNA clone E10379, mRNA sequence.
ACCESSION C19416 C99186
VERSION C19416.2 GI:15072238
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 440)
AUTHORS Sasaki,T.
TITLE Rice cDNA from panicle at ripening stage
JOURNAL Unpublished (1996)
COMMENT On Oct 24, 1996 this sequence version replaced gi:1631687 gi:3761938.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT ='RGP'.
FEATURES Location/Qualifiers
source 1. .440
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="E10379"
/clone_lib="Rice panicle at ripening stage"
/dev_stage="ripening stage"
/note="Organ: panicle; Rice cDNA from panicle at ripening stage"
BASE COUNT 60 a 152 c 159 g 68 t 1 others
ORIGIN

Query Match 29.9%; Score 30.2; DB 14; Length 440;
Best Local Similarity 58.2%; Pred. No. 1.4e+02;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy       6 GGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGC 65
           |||||||  || | | | || || || |  || ||    | |||  |||     |  |||
Db     131 GGTGCTGGTCATCTTCAGCCCCGTGCTGGTGCCGGCGGCCATCGCGCTGGCGCTCATGGC 190

Qy      66 GTCGGGCGTCTTCGGCTCGATCGGCCAGGGC 96
           | | ||  || ||  ||||  ||| |  |||
Db     191 GGCCGGGTTCGTCACCTCGGGCGGGCTCGGC 221
```

RESULT 10
BQ172791
LOCUS BQ172791 571 bp mRNA linear EST 29-APR-2002
DEFINITION 1008b05 rice seed cDNA library Oryza sativa cDNA, mRNA sequence.
ACCESSION BQ172791
VERSION BQ172791.1 GI:20337161
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 571)
AUTHORS Wang,X., Zhao,J., Li,Q. and Zhang,X.
TITLE Expressed sequence tags analysis of a rice seed cDNA library
JOURNAL Unpublished (2002)
COMMENT Contact: Xiansheng Zhang
Plant development molecular biology lab
college of life sciences
Shan dong agriculture university, Tai'an, Shan dong, China
Tel: 086-0538-8241144
Email: zhiwufy@sdau.edu.cn.
FEATURES Location/Qualifiers
source 1. .571
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone_lib="rice seed cDNA library"
/tissue_type="seed"
/note="Vector: pT-Adv"
BASE COUNT 83 a 189 c 198 g 101 t
ORIGIN

Query Match 29.9%; Score 30.2; DB 14; Length 571;
Best Local Similarity 58.2%; Pred. No. 1.5e+02;
Matches 53; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy       6 GGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGC 65
           |||||||  || | | | || || || |  || ||    | |||  |||     |  |||
Db     293 GGTGCTGGTCATCTTCAGCCCCGTGCTGGTGCCGGCGGCCATCGCGCTGGCGCTCATGGC 352

Qy      66 GTCGGGCGTCTTCGGCTCGATCGGCCAGGGC 96
           | | ||  || ||  ||||  ||| |  |||
Db     353 GGCCGGGTTCGTCACCTCGGGCGGGCTCGGC 383
```

RESULT 11
BM920552/c
LOCUS BM920552 1146 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6709522 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5750558 5', mRNA sequence.
ACCESSION BM920552
VERSION BM920552.1 GI:19370931
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1146)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12781 row: 1 column: 15

```
          High quality sequence stop: 539.
FEATURES             Location/Qualifiers
     source          1. .1146
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5750558"
                     /clone_lib="NIH_MGC_122"
                     /lab_host="DH10B"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
BASE COUNT      274 a    326 c    298 g    248 t
ORIGIN

  Query Match             29.9%;  Score 30.2;  DB 14;  Length 1146;
  Best Local Similarity   58.2%;  Pred. No. 1.8e+02;
  Matches   53;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       3 ATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTC 62
           | ||||  | |   ||||  |||  ||    || ||||||||||||  ||   ||| || |
Db     818 AAGGGTTTTTCTTCCCCTGCCCCTTGAAATCCCCCTGGACCTCTTTTTAGACTTTCCCCC 759

Qy      63 GGCGTCGGGCGTCTTCGGCTCGATCGGCCAG 93
             | |||  | || || |||    ||| |
Db     758 CATCTGGGGTCTTTTGGGTTCGGGTTGCCTG 728


RESULT 12
BE495030/c
LOCUS       BE495030                455 bp    mRNA    linear   EST 02-AUG-2000
DEFINITION  WHE1267_F04_K07ZS Secale cereale anther cDNA library Secale cereale
            cDNA clone WHE1267_F04_K07, mRNA sequence.
ACCESSION   BE495030
VERSION     BE495030.1  GI:9661623
KEYWORDS    EST.
SOURCE      rye.
  ORGANISM  Secale cereale
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Secale.
REFERENCE   1  (bases 1 to 455)
  AUTHORS   Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton
            ,R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,
            Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Anther cDNA library from rye
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .455
                     /organism="Secale cereale"
                     /cultivar="Blanco"
                     /db_xref="taxon:4550"
                     /clone="WHE1267_F04_K07"
                     /clone_lib="Secale cereale anther cDNA library"
                     /tissue_type="Anther"
                     /dev_stage="Adult plant before anthesis"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Anthers were harvested and pooled  from early
                     meiosis to late meiosis.  The tissue, total RNA, and
                     poly(A)  RNA were prepared (Butler, Ross and Gustafson) at
                     University of  Missouri, Columbia.  A cDNA library was
                     made, and the cDNA clones were  in vivo excised to give
                     pBluescript phagemids in the TJ Close lab  (Choi, Close,
                     Fenton) at the University of California, Riverside.
                     Plasmid DNA preparations and DNA sequencing were performed
                     in the OD  Anderson lab (all other authors)."
BASE COUNT       79 a    112 c    144 g    120 t
ORIGIN

  Query Match             29.7%;  Score 30;  DB 10;  Length 455;
  Best Local Similarity   57.4%;  Pred. No. 1.6e+02;
  Matches   54;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           | | |  ||| ||||   | | |||||   || ||||| |||  |   |    | | |||
Db     131 GGACGACTGCGGCACGTACATCCCCGCCGCCACCCGCTAGACGACGGTGATGAGCTCTCC 72

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGG 94
           ||  | |   ||  || || |||   |||||| |
Db      71 TCCTCCTGCTGCTGCTGCGACTCCGGCGGCCACG 38


RESULT 13
BI152354/c
LOCUS       BI152354                858 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602917523F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:5068204 5',
            mRNA sequence.
ACCESSION   BI152354
VERSION     BI152354.1  GI:14612355
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 858)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11183  row: m  column: 05
            High quality sequence start: 17
            High quality sequence stop: 654.
FEATURES             Location/Qualifiers
     source          1. .858
                     /organism="Mus musculus"
                     /strain="CZECH II"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5068204"
                     /clone_lib="NCI_CGAP_Lu29"
                     /tissue_type="spontaneous tumor, metastatic to mammary.
                     Stem cell origin."
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      191 a    243 c    252 g    172 t
ORIGIN
```

```
  Query Match             29.7%;  Score 30;  DB 13;  Length 858;
  Best Local Similarity   67.7%;  Pred. No. 1.9e+02;
  Matches   42;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy      37 CTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGC 96
           | ||| ||| ||||||| | ||| ||   ||   | ||||| || | | ||| || |||
Db     850 CCGGATCTCATCGTACTTTCTTCTTCTCAGTTCAGTGTCTTTGGTTAGGTCGACCGGGGG 791

Qy      97 AG 98
           ||
Db     790 AG 789


RESULT 14
BM452552/c
LOCUS       BM452552                988 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6401029 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493399
            5', mRNA sequence.
ACCESSION   BM452552
VERSION     BM452552.1  GI:18501592
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12117  row: a  column: 16
            High quality sequence stop: 504.
FEATURES             Location/Qualifiers
     source          1. .988
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5493399"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      262 a    242 c    296 g    182 t      6 others
ORIGIN

  Query Match             29.7%;  Score 30;  DB 13;  Length 988;
  Best Local Similarity   64.3%;  Pred. No. 2e+02;
  Matches   45;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      31 CAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGC 90
           |||| || ||||| |    ||| |||| ||  |   |||| || ||| ||| ||| || |
Db     165 CAGCTGCAGGACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCTTTGGCGCGAGCGCC 106

Qy      91 CAGGGCAGCT 100
           |   ||  ||
Db     105 CGTCGCTCCT 96


RESULT 15
AA081455/c
LOCUS       AA081455                371 bp    mRNA    linear   EST 23-DEC-1997
DEFINITION  zn04f10.s1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
            IMAGE:546475 3', mRNA sequence.
ACCESSION   AA081455
VERSION     AA081455.1  GI:1623513
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chissoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            ,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            ,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Trevaskis,E.,
            Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M.
  TITLE     Generation and analysis of 280,000 human expressed sequence tags
  JOURNAL   Genome Res. 6 (9), 807-828 (1996)
  MEDLINE   97044478
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            WARNING: There is evidence that suggests that the 384-well parent
            plate of this clone contains both human and mouse derived clones.
            Thus, the origin of this clone is uncertain. This caution should be
            kept in mind should you use this clone.

            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40M13 fwd. from Amersham.
FEATURES             Location/Qualifiers
     source          1. .371
                     /organism="Homo sapiens"
                     /db_xref="GDB:3924665"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:546475"
                     /clone_lib="Stratagene hNT neuron (#937233)"
                     /dev_stage="hNT neurons"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI; Cloned unidirectionally.  Primer: Oligo dT.
                     Differentiated, post mitotic hNT neurons. Average insert
                     size: 1.5 kb; Uni-ZAP XR Vector; ~5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3'"
BASE COUNT      125 a     68 c     81 g     97 t
ORIGIN

  Query Match             29.5%;  Score 29.8;  DB 9;  Length 371;
  Best Local Similarity   70.2%;  Pred. No. 1.7e+02;
  Matches   40;  Conservative    0;  Mismatches   17;  Indels    0;  Gaps    0;

Qy       2 CATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTT 58
           || || |||||||| | | |||| ||  ||  || |  | |||| ||  ||||||||
Db     360 CAAGGTTGCTGCACCCTCATACCAGCTGGCTCCCTCCTGCCCTCCTCTGACTGTTTT 304


Search completed: November  5, 2002, 16:17:40
Job time : 1119.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050
Perfect score: 101
Sequence: 1 gcatgggtgctgcacaccct.....tcgatcggccagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	2	101	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	3	101	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	4	101	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	5	101	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
	6	101	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
	7	101	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
	8	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	9	49.8	49.3	15872	4	US-09-105-537-1	Sequence 1, Appli
c	10	41.8	41.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	11	41.8	41.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	12	40.8	40.4	33529	4	US-09-144-085-3	Sequence 3, Appli
	13	40.8	40.4	80161	3	US-09-036-987A-1	Sequence 1, Appli
	14	40.8	40.4	80161	4	US-09-370-700-1	Sequence 1, Appli
	15	40.2	39.8	4689	4	US-09-105-537-34	Sequence 34, Appl
	16	40.2	39.8	13842	4	US-09-105-537-30	Sequence 30, Appl
	17	40.2	39.8	36778	4	US-09-105-537-5	Sequence 5, Appli
	18	40.2	39.8	38506	3	US-09-320-878-19	Sequence 19, Appl
	19	37.6	37.2	11219	1	US-07-642-734C-1	Sequence 1, Appli
	20	37.6	37.2	11219	3	US-08-439-009A-1	Sequence 1, Appli
	21	36.2	35.8	13987	2	US-08-804-227C-13	Sequence 13, Appl
	22	36.2	35.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
	23	36.2	35.8	44377	2	US-08-804-198-1	Sequence 1, Appli
	24	35.4	35.0	28958	1	US-08-258-261B-6	Sequence 6, Appli
	25	35.4	35.0	28958	1	US-08-456-837-6	Sequence 6, Appli
	26	35.4	35.0	28958	1	US-08-457-342-6	Sequence 6, Appli
	27	35.4	35.0	28958	1	US-08-457-646A-6	Sequence 6, Appli
	28	35.4	35.0	28958	1	US-08-458-076A-6	Sequence 6, Appli
	29	35.4	35.0	28958	1	US-08-764-233A-4	Sequence 4, Appli
	30	35.4	35.0	28958	1	US-08-457-335A-6	Sequence 6, Appli
	31	35.4	35.0	28958	1	US-08-729-214-6	Sequence 6, Appli
	32	35.4	35.0	28958	3	US-09-028-934-6	Sequence 6, Appli
	33	35.4	35.0	49377	1	US-08-764-233A-1	Sequence 1, Appli
	34	34.4	34.1	43280	2	US-08-804-227C-1	Sequence 1, Appli
	35	34	33.7	1681	4	US-09-434-288-7	Sequence 7, Appli
	36	33	32.7	50937	4	US-09-428-517-1	Sequence 1, Appli
	37	31.4	31.1	11220	4	US-09-105-537-32	Sequence 32, Appl
	38	31.2	30.9	477	3	US-09-010-809-2	Sequence 2, Appli
c	39	31.2	30.9	477	3	US-09-010-809-17	Sequence 17, Appl
	40	31	30.7	1601	1	US-08-206-006-1	Sequence 1, Appli
	41	30.6	30.3	20235	1	US-07-642-734C-3	Sequence 3, Appli
	42	30.6	30.3	20235	3	US-08-439-009A-3	Sequence 3, Appli
	43	30.4	30.1	11561	1	US-08-450-332-1	Sequence 1, Appli
	44	30.4	30.1	11561	2	US-08-637-640-1	Sequence 1, Appli
	45	30.4	30.1	11561	4	US-09-004-406C-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 101; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.2e-21;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 101;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 101;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19950 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 20009

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  20010 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 20050


RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             49.3%;  Score 49.8;  DB 4;  Length 15872;
  Best Local Similarity   68.3%;  Pred. No. 2.2e-06;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || |||   ||||||   || |||   || | || ||| |||  |||||| || || |||
Db   7682 GCGTGGAACCTGCACGAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCC 7741

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          || |  || ||| || ||||| |   |||||||| || |||
```

```
Db   7742 TCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTA 7782


RESULT 10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             41.4%;  Score 41.8;  DB 4;  Length 4403765;
  Best Local Similarity   63.4%;  Pred. No. 0.0013;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||| | ||||       ||| |||  |||||||| |||   |||    |||||||
Db 4286204 GCCTGGGAGTTGCATGAAGCAACCAGCGGACAGCCGCTTGACTGGTTCTGCTTGTTTTCG 4286145

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||||| | |   ||  |||||||   || ||| |   |||
Db 4286144 TCGGCGGCAGCTCTCACCGGCTCGCCGGGACAGAGTGCCTA 4286104


RESULT 11
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             41.4%;  Score 41.8;  DB 4;  Length 4411529;
  Best Local Similarity   63.4%;  Pred. No. 0.0013;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||| | ||||       ||| |||  |||||||| |||   |||    |||||||
Db 4293947 GCCTGGGAGTTGCATGAAGCAACCAGCGGACAGCCGCTTGACTGGTTCTGCTTGTTTTCG 4293888

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||||| | |   ||  |||||||   || ||| |   |||
Db 4293887 TCGGCGGCAGCTCTCACCGGCTCGCCGGGACAGAGTGCCTA 4293847


RESULT 12
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             40.4%;  Score 40.8;  DB 4;  Length 33529;
  Best Local Similarity   65.2%;  Pred. No. 0.0013;
  Matches   60;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||||||   || || || |||| |   || |     ||||| ||||| || ||||||
Db  19898 CTGCACGAGCTGACGCGGGAGCTGGATCTCGCGGCGTTCGTGCTGTTCTCGTCGGCGGCC 19957

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           ||    || ||| ||   ||||| ||| |||
Db  19958 GGTACGTTTGGCGCGGCGGGCCAGAGCAACTA 19989


RESULT 13
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
```

```
;       FILING DATE:  09-MAR-1998
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Stuart, Donald R
;       REGISTRATION NUMBER:  28,479
;       REFERENCE/DOCKET NUMBER:  50,608
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (317)337-4816
;       TELEFAX:  (317)337-4847
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  80161 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             40.4%;  Score 40.8;  DB 3;  Length 80161;
  Best Local Similarity   65.2%;  Pred. No. 0.0014;
  Matches   60;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||| ||   ||     ||| ||| ||| ||||| |||||| ||||| |||||   |||
Db  58799 CTGGACCGGCTGCTTGGCGGGCAACCGTTGGACGCCTTCGTGCTGTTCTCCTCGATCTCG 58858

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || || |  ||  |   |||||| ||   |||
Db  58859 GGAGTTTGGGGAGCCGGCGGCCAAGGAGCCTA 58890


RESULT 14
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             40.4%;  Score 40.8;  DB 4;  Length 80161;
  Best Local Similarity   65.2%;  Pred. No. 0.0014;
  Matches   60;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||| ||   ||     ||| ||| ||| ||||| |||||| ||||| |||||   |||
Db  58799 CTGGACCGGCTGCTTGGCGGGCAACCGTTGGACGCCTTCGTGCTGTTCTCCTCGATCTCG 58858

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || || |  ||  |   |||||| ||   |||
Db  58859 GGAGTTTGGGGAGCCGGCGGCCAAGGAGCCTA 58890


RESULT 15
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-34

  Query Match             39.8%;  Score 40.2;  DB 4;  Length 4689;
  Best Local Similarity   62.4%;  Pred. No. 0.0014;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           ||   ||| ||  ||  |||   |||||     |||||||| ||||||  | |  |||
Db    3688 GCCGAGGTCCTCGACGACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCC 3747

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||    | || ||||  |||     | ||||||||  |||
Db    3748 TCGAACGCCGGGGTCTGGGGCAGCGGCAGCCAGGGCGTCTA 3788


Search completed: November  5, 2002, 14:05:11
Job time : 1655.74 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050
Perfect score: 101
Sequence: 1 gcatgggtgctgcacaccct.....tcgatcggccagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	49.8	49.3	15872	10	US-09-861-289-1	Sequence 1, Appli
	2	40.2	39.8	4689	10	US-09-861-289-34	Sequence 34, Appl
	3	40.2	39.8	13842	10	US-09-861-289-30	Sequence 30, Appl
	4	40.2	39.8	36778	10	US-09-861-289-5	Sequence 5, Appli
	5	31.4	31.1	11220	10	US-09-861-289-32	Sequence 32, Appl
	6	28.6	28.3	1059	10	US-09-815-242-7877	Sequence 7877, Ap
	7	28.4	28.1	299	10	US-09-294-093B-3349	Sequence 3349, Ap
c	8	28	27.7	219	10	US-09-777-564-12	Sequence 12, Appl
c	9	28	27.7	219	10	US-09-777-564-1473	Sequence 1473, Ap
	10	27.6	27.3	518	10	US-09-764-887-133	Sequence 133, App
	11	27.6	27.3	518	10	US-09-764-887-603	Sequence 603, App
	12	27.6	27.3	518	10	US-09-764-887-604	Sequence 604, App
	13	27.6	27.3	519	10	US-09-764-887-605	Sequence 605, App
c	14	27.4	27.1	204	10	US-09-983-965-5085	Sequence 5085, Ap
	15	27.2	26.9	23668	10	US-09-741-148A-3	Sequence 3, Appli
	16	26.8	26.5	1128	10	US-09-815-242-7706	Sequence 7706, Ap
	17	26.6	26.3	1566	10	US-09-815-242-7758	Sequence 7758, Ap
	18	26.6	26.3	3396	10	US-09-749-728B-32	Sequence 32, Appl
	19	26.6	26.3	4015	10	US-09-733-294A-3	Sequence 3, Appli
	20	26.6	26.3	4015	10	US-09-990-080-1	Sequence 1, Appli
	21	26.6	26.3	5011	10	US-09-764-847-1562	Sequence 1562, Ap
c	22	26.4	26.1	190	10	US-09-920-300A-103	Sequence 103, App
c	23	26.4	26.1	190	12	US-10-033-528-103	Sequence 103, App
	24	26.4	26.1	234	10	US-09-777-564-397	Sequence 397, App
	25	26.4	26.1	234	10	US-09-777-564-654	Sequence 654, App
	26	26.4	26.1	234	10	US-09-777-564-1430	Sequence 1430, Ap
	27	26.4	26.1	234	10	US-09-777-564-1464	Sequence 1464, Ap
	28	26.4	26.1	234	10	US-09-864-864-143	Sequence 143, App
c	29	26.4	26.1	234	10	US-09-920-300A-302	Sequence 302, App
c	30	26.4	26.1	234	10	US-09-920-300A-1026	Sequence 1026, Ap
	31	26.4	26.1	234	10	US-09-920-300A-1489	Sequence 1489, Ap
c	32	26.4	26.1	234	10	US-09-920-300A-1646	Sequence 1646, Ap
c	33	26.4	26.1	234	10	US-09-998-598-1646	Sequence 1646, Ap
c	34	26.4	26.1	234	12	US-10-033-528-302	Sequence 302, App
c	35	26.4	26.1	234	12	US-10-033-528-1026	Sequence 1026, Ap
	36	26.4	26.1	234	12	US-10-033-528-1489	Sequence 1489, Ap
c	37	26.4	26.1	234	12	US-10-033-528-1646	Sequence 1646, Ap
c	38	26.4	26.1	312	10	US-09-920-300A-75	Sequence 75, Appl
c	39	26.4	26.1	312	12	US-10-033-528-75	Sequence 75, Appl
c	40	26.4	26.1	316	10	US-09-920-300A-67	Sequence 67, Appl
c	41	26.4	26.1	316	12	US-10-033-528-67	Sequence 67, Appl
c	42	26.4	26.1	359	10	US-09-777-564-556	Sequence 556, App
c	43	26.4	26.1	359	10	US-09-777-564-1432	Sequence 1432, Ap
	44	26.4	26.1	634	10	US-09-922-217-868	Sequence 868, App
	45	26.4	26.1	634	10	US-09-833-263-868	Sequence 868, App

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 49.3%; Score 49.8; DB 10; Length 15872;
Best Local Similarity 68.3%; Pred. No. 1e-06;
Matches 69; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||   || |||   || | || ||| |||  |||||| || || |||
Db    7682 GCGTGGAACCTGCACGAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCC 7741

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
             || |  || ||| || ||||| |   |||||||| || |||
Db    7742 TCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTA 7782
```

RESULT 2
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.

```
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             39.8%;  Score 40.2;  DB 10;  Length 4689;
  Best Local Similarity   62.4%;  Pred. No. 0.00078;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           ||   ||| ||  ||  |||   |||||     |||||||||  |||||| || |  |||
Db    3688 GCCGAGGTCCTCGACGACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCC 3747

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||    | || ||||  |||     | | ||||||  |||
Db    3748 TCGAACGCCGGGGTCTGGGGCAGCGGCAGCCAGGGCGTCTA 3788


RESULT 3
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             39.8%;  Score 40.2;  DB 10;  Length 13842;
  Best Local Similarity   62.4%;  Pred. No. 0.00088;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || | ||||||  ||   || ||||| || |     || |||   ||||| || || ||
Db    6964 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 7023

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || | |||| ||    | |||     ||| |||||||  |||
Db    7024 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTA 7064


RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             39.8%;  Score 40.2;  DB 10;  Length 36778;
  Best Local Similarity   62.4%;  Pred. No. 0.00099;
  Matches   63;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || | ||||||  ||   || ||||| || |     || |||   ||||| || || ||
Db    8705 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 8764

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || | |||| ||    | |||     ||| |||||||  |||
Db    8765 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTA 8805


RESULT 5
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             31.1%;  Score 31.4;  DB 10;  Length 11220;
  Best Local Similarity   59.6%;  Pred. No. 0.42;
  Matches   53;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy      13 CACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGC 72
           | ||||   || | ||   |    ||||     |||||  |||| ||||| ||  | | |
Db  10192 CTCACCTCGACGCCCGGCTACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCCGCC 10251

Qy      73 GTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           ||||||||       || ||||||  |||
Db  10252 GTCTTCGGTGGCGCGGGGCAGGGCGCCTA 10280


RESULT 6
US-09-815-242-7877
; Sequence 7877, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
```

```
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7877
;   LENGTH: 1059
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1059)
US-09-815-242-7877

  Query Match            28.3%;  Score 28.6;  DB 10;  Length 1059;
  Best Local Similarity  57.1%;  Pred. No. 2.3;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       6 GGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGC 65
           |||||||   | ||| |   |||||     |||||||  |||||  ||| |   |   |
Db     792 GGTGCTGGTGAGCCTGATGGGCGAGGTCAGGCTGGACGCCTTCGACCTGGTCAACCGCGA 851

Qy      66 GTCGGGCGTCTTCGGCTCGATCGGCCAGGGC 96
             | || |  |||||    |||||  |  ||
Db     852 ACTGCGCCTGCTCGGCAGCGTCGGCTACCGC 882


RESULT 7
US-09-294-093B-3349
; Sequence 3349, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
;  APPLICANT:      Lalgudi, Raghunath, V.
;  APPLICANT:  Ito, Laura, Y.
;  APPLICANT:  Sherman, Bradley, K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
;  FILE REFERENCE:  PL-0009 US
;  CURRENT APPLICATION NUMBER: US/09/294,093B
;  CURRENT FILING DATE:  1999-04-16
;  PRIOR APPLICATION NUMBER:  60/082,567
;  PRIOR FILING DATE:  April 21, 1998
;  NUMBER OF SEQ ID NOS:      6207
;  SOFTWARE:      PERL Program
; SEQ ID NO 3349
;   LENGTH: 299
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20010051335A1 700379878H1
;   NAME/KEY: unsure
;   LOCATION: 60, 214, 278
;   OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-3349

  Query Match            28.1%;  Score 28.4;  DB 10;  Length 299;
  Best Local Similarity  59.5%;  Pred. No. 2.3;
  Matches   47;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      18 CCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTT 77
           | | || |||||   |||     |  ||| |||   | | | |||| |||| |||||| |
Db     215 CTTCACGCGCGACGGGCCCGCCCAGTTCTCCGTGGCGGTGTTCTCGTCGTCAGGCGTCAT 274

Qy      78 CGGCTCGATCGGCCAGGGC 96
           ||   ||   | || || |
Db     275 CGANCCGGCAGCCCTGGAC 293


RESULT 8
US-09-777-564-12/c
; Sequence 12, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE:  210121.493
;  CURRENT APPLICATION NUMBER: US/09/777,564
;  CURRENT FILING DATE:  2001-02-05
;  NUMBER OF SEQ ID NOS:  1730
;  SOFTWARE:  FastSEQ for Window Version 4.0
; SEQ ID NO 12
;   LENGTH: 219
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-777-564-12

  Query Match            27.7%;  Score 28;  DB 10;  Length 219;
  Best Local Similarity  60.5%;  Pred. No. 2.9;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      15 CACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGT 74
           ||| || |||||   |   |   |  ||||  ||||||| ||||||| | | | | |
Db      88 CACACTGACCCGGCCTCCCACTTGGCCCCTCAGCGTACTGATTTCCTCCTCATGGTTCTT 29

Qy      75 CTTCGGCTCGATCGGC 90
           |||| | | |  | ||
Db      28 CTTCAGGTAGGCCAGC 13


RESULT 9
US-09-777-564-1473/c
; Sequence 1473, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
;  APPLICANT:  Algate, Paul A.
;  APPLICANT:  Mannion, Jane
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
;  FILE REFERENCE:  210121.493
;  CURRENT APPLICATION NUMBER: US/09/777,564
;  CURRENT FILING DATE:  2001-02-05
;  NUMBER OF SEQ ID NOS:  1730
;  SOFTWARE:  FastSEQ for Window Version 4.0
; SEQ ID NO 1473
;   LENGTH: 219
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-777-564-1473

  Query Match            27.7%;  Score 28;  DB 10;  Length 219;
  Best Local Similarity  60.5%;  Pred. No. 2.9;
  Matches   46;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;
```

```
Qy      15 CACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGT 74
           ||| || |||||    |   |   |  ||||  ||||||| |||||||  | | |  | |
Db      88 CACACTGACCCGGCCTCCCACTTGGCCCCTCAGCGTACTGATTTCCTCCTCATGGTTCTT 29

Qy      75 CTTCGGCTCGATCGGC 90
           |||| | | |  | ||
Db      28 CTTCAGGTAGGCCAGC 13


RESULT 10
US-09-764-887-133
; Sequence 133, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 133
;   LENGTH: 518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-133

  Query Match             27.3%;  Score 27.6;  DB 10;  Length 518;
  Best Local Similarity   63.6%;  Pred. No. 4.3;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      35 CGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGG 94
           ||| ||||| |    ||| |||| ||  |   |||| || ||| ||| ||| || ||
Db     258 CGCAGGACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCTTTGGCGCGAGCGCCCGTC 317

Qy      95 GCAGCT 100
           ||  ||
Db     318 GCTCCT 323


RESULT 11
US-09-764-887-603
; Sequence 603, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 603
;   LENGTH: 518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-603

  Query Match             27.3%;  Score 27.6;  DB 10;  Length 518;
  Best Local Similarity   63.6%;  Pred. No. 4.3;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      35 CGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGG 94
           ||| ||||| |    ||| |||| ||  |   |||| || ||| ||| ||| || ||
Db     258 CGCAGGACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCTTTGGCGCGAGCGCCCGTC 317

Qy      95 GCAGCT 100
           ||  ||
Db     318 GCTCCT 323


RESULT 12
US-09-764-887-604
; Sequence 604, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 604
;   LENGTH: 518
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-604

  Query Match             27.3%;  Score 27.6;  DB 10;  Length 518;
  Best Local Similarity   63.6%;  Pred. No. 4.3;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      35 CGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGG 94
           ||| ||||| |    ||| |||| ||  |   |||| || ||| ||| ||| || ||
Db     258 CGCAGGACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCTTTGGCGCGAGCGCCCGTC 317

Qy      95 GCAGCT 100
           ||  ||
Db     318 GCTCCT 323


RESULT 13
US-09-764-887-605
; Sequence 605, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 605
;   LENGTH: 519
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-605

  Query Match             27.3%;  Score 27.6;  DB 10;  Length 519;
  Best Local Similarity   63.6%;  Pred. No. 4.3;
  Matches   42;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy      35 CGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGG 94
           ||| ||||| |    ||| |||| ||  |   |||| || ||| ||| ||| || ||
Db     259 CGCAGGACCGCGCTTTACAGTTTACCAGGCACTCGGACGCCTTGGGCGCGAGCGCCCGTC 318

Qy      95 GCAGCT 100
           ||  ||
Db     319 GCTCCT 324


RESULT 14
US-09-983-965-5085/c
; Sequence 5085, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
```

```
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE:  2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 5085
;   LENGTH: 204
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   FEATURE:
;   OTHER INFORMATION: Clone ID: 31-LIB34-062-Q1-E1-H3
US-09-983-965-5085

  Query Match             27.1%;  Score 27.4;  DB 10;  Length 204;
  Best Local Similarity   54.5%;  Pred. No. 4.4;
  Matches   55;  Conservative    0;  Mismatches   46;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || ||  | |||      || |  ||||||||||||  |  ||      | | |||  ||
Db     141 GCGTGTCTCCTGTCACGTCTGAGGCGCGAGCAGCCGGGGACCCAGCAGCTCCCGTTCGCC 82

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           | |  |||     |  |  | || || |||||  |||| |
Db      81 TGGCGGTCTTAGCGCGCCCCCGCGCTCCGCCAGACCAGCGA 41


RESULT 15
US-09-741-148A-3
; Sequence 3, Application US/09741148A
; Patent No. US20020076750A1
; GENERAL INFORMATION:
;  APPLICANT: Chunhua YAN et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000566
;  CURRENT APPLICATION NUMBER: US/09/741,148A
;  CURRENT FILING DATE:  2002-04-15
;  PRIOR APPLICATION NUMBER: 60/206,982
;  PRIOR FILING DATE: 2000-05-25
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 23668
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(23668)
;   OTHER INFORMATION: n = A,T,C or G
US-09-741-148A-3

  Query Match             26.9%;  Score 27.2;  DB 10;  Length 23668;
  Best Local Similarity   56.8%;  Pred. No. 8.9;
  Matches   50;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      13 CACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCGGGC 72
           ||| ||  | |||  | |||| |   ||   |   | | |||   |||   ||  | | |
Db  13092 CACTCCTGTCCCCAGGTGCAGGCCACGGTGGTGGGCTTCCTGGCGTCCATCGCAGCCGTC 13151

Qy      73 GTCTTCGGCTCGATCGGCCAGGGCAGCT 100
           ||||| |||| ||||    | |||  ||
Db  13152 GTCTTTGGCTGGATCCCTGATGGCCACT 13179


Search completed: November  5, 2002, 22:59:34
Job time : 99.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050
Perfect score: 101
Sequence: 1 gcatgggtgctgcacaccct.....tcgatcggccagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	101	100.0	71989	21	AAA29349	Sorangium cellulos
	3	49.8	49.3	15872	21	AAZ87283	S. venezuelae vep
	4	48.2	47.7	125401	22	AAD17186	Streptomyces nours
	5	46.6	46.1	552	24	AAI68063	Epothilone PKS KR
	6	46.6	46.1	6462	22	AAF90037	Nucleotide sequenc
	7	46.6	46.1	15872	18	AAT68715	Streptomyces venez
	8	46.6	46.1	34071	22	AAF90033	Nucleotide sequenc
c	9	46.6	46.1	42717	22	AAF90032	Nucleotide sequenc
	10	45.6	45.1	552	24	AAI68064	9-keto-epothilone
	11	45.6	45.1	5676	19	AAV21186	Amycolatopsis medi
	12	45.6	45.1	27541	22	AAD17185	Streptomyces nours
	13	45.6	45.1	53789	19	AAV21187	Amycolatopsis medi
	14	44	43.6	6333	22	AAF31641	Mycobacterium tube
	15	44	43.6	8301	22	AAF90035	Nucleotide sequenc
	16	43.4	43.0	20394	22	AAF24892	Pimaricin biosynth
	17	43.4	43.0	30690	21	AAA92301	S. avermitilis ave
	18	43.4	43.0	30690	22	AAH79277	Streptomyces averm
	19	43.4	43.0	65140	22	AAD17184	Streptomyces nours
	20	43.4	43.0	77536	21	AAA14651	Nucleotide sequenc
	21	41.8	41.4	47981	22	AAF30757	Micromonospora meg
c	22	41.8	41.4	4403765	22	AAI99683	Mycobacterium tube
c	23	41.8	41.4	4411529	22	AAI99682	Mycobacterium tube
	24	41.2	40.8	11916	22	AAH79279	Streptomyces averm
	25	41.2	40.8	12381	21	AAZ58381	Streptomyces averm
	26	41.2	40.8	31422	21	AAA92302	S. avermitilis ave
	27	41.2	40.8	31422	22	AAH79278	Streptomyces averm
	28	40.8	40.4	14775	22	AAF88338	S. spinosa DNA fra
	29	40.8	40.4	33529	23	AAS17367	DNA sequence of S.
	30	40.8	40.4	50000	22	AAF88313	S. spinosa DNA fra
	31	40.8	40.4	50000	22	AAF88316	S. spinosa DNA fra
	32	40.8	40.4	80161	20	AAZ21501	DNA fragment of Sa
	33	40.2	39.8	4466	21	AAA14663	Nucleotide sequenc
	34	40.2	39.8	4478	21	AAA14661	Nucleotide sequenc
	35	40.2	39.8	4547	21	AAA14664	Nucleotide sequenc
	36	40.2	39.8	4571	21	AAA14662	Nucleotide sequenc
	37	40.2	39.8	4689	21	AAZ87299	S. venezuelae macr
	38	40.2	39.8	5088	22	AAF90038	Nucleotide sequenc
	39	40.2	39.8	5292	22	AAF90036	Nucleotide sequenc
	40	40.2	39.8	13842	21	AAZ87297	S. venezuelae macr
	41	40.2	39.8	36778	21	AAZ87318	S. venezuelae pik
	42	40.2	39.8	37948	21	AAZ87285	S. venezuelae pik
	43	40.2	39.8	38506	21	AAA75633	Nucleotide sequenc
	44	40.2	39.8	38506	21	AAZ56001	Recombinant cosmid
c	45	40.2	39.8	77536	21	AAA14651	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;    99WO-EP04171.
XX
PR   18-JUN-1998;    98US-0099504.
PR   24-SEP-1998;    98US-0101631.
PR   05-FEB-1999;    99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 101;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-21;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19950 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 20009

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  20010 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 20050


RESULT 3
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO 1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag= b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag= c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match             49.3%;  Score 49.8;  DB 21;  Length 15872;
  Best Local Similarity   68.3%;  Pred. No. 1.4e-05;
  Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || |||   ||||||   || |||   || | || ||| |||  |||||| || || |||
Db   7682 GCGTGGAACCTGCACGAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCC 7741

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          || |  || ||| || ||||| |   ||||||| || |||
Db   7742 TCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTA 7782
```

```
RESULT 4
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             47.7%;  Score 48.2;  DB 22;  Length 125401;
  Best Local Similarity   67.3%;  Pred. No. 5.5e-05;
  Matches   68;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||  |   |||||||  |     |||| |  |||||| ||||  |||
Db   55745 GCCTGGCATCTGCACGACGCCACCCGCGGCCTCGACCTGGCCGGCTTCGTCCTGTACTCC 55804

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
             ||||  || |||||| | |||     ||||||||||| |||
Db   55805 TCGGTCTCCGGCGTCCTGGGCAGCCCCGGCCAGGGCAACTA 55845


RESULT 5
AAI68063
ID   AAI68063 standard; DNA; 552 BP.
XX
AC   AAI68063;
XX
DT   13-MAR-2002  (first entry)
XX
DE   Epothilone PKS KR domain nucleotide sequence.
XX
KW   Cystobacterineae; recombinant; polyketide synthase; PKS; polyketide;
KW   epothilone; pilA gene; cytostatic; antipsoriatic; antiarthritic;
KW   antiarteriosclerotic; antiinflammatory; neuroprotective; vasotropic; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..552
FT                   /*tag= a
FT                   /note= "KR domain of extender module 6"
XX
PN   WO200183800-A2.
XX
PD   08-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13793.
XX
PR   28-APR-2000; 2000US-0560367.
PR   14-SEP-2000; 2000US-232696P.
PR   21-DEC-2000; 2000US-257517P.
PR   03-APR-2001; 2001US-0825856.
PR   03-APR-2001; 2001US-0825876.
PR   13-APR-2001; 2001US-269020P.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Arslanian RL,  Ashley G,  Frykman S,  Julien B,  Katz L,  Khosla C;
PI   Lau J,  Licardi PJ,  Regentin R,  Santi D,  Tang L;
XX
DR   WPI; 2002-075167/10.
DR   P-PSDB; ABB07167.
XX
PT   Recombinant host cells useful for producing polyketides e.g. epothilone
PT   or its derivatives, comprises a recombinant expression vector encoding
PT   a heterologous polyketide synthase gene  -
XX
PS   Example 11; Page 164-165; 221pp; English.
```

```
XX
CC   The invention provides a recombinant host cell, of the suborder
CC   Cystobacterineae comprising a recombinant expression vector encoding a
CC   heterologous polyketide synthase (PKS) gene and produces a polyketide
CC   synthesized by the PKS enzyme encoded on the vector. An epothilone
CC   derivative of a specified formula can be produced by culturing the host
CC   cell with a diketide equivalent compound of a specified formula. The host
CC   cells produces polyketides at high levels and are used in the production
CC   of not only epothilones, including new epothilone derivatives, but also
CC   other polyketides. Methods of purifying the epothilone derivatives are
CC   also useful for treating cancer, hyperproliferative diseases and
CC   conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC   lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC   /or restenosis. It improves polyketide production in any organism and
CC   also for production of products of recombinant PKS genes and modification
CC   enzymes. The present sequence represents the nucleotide sequence of the
CC   KR domain of extender module 6 of the epothilone PKS. Inactivation of
CC   this domain results in a novel PKS capable of producing a 9-keto-
CC   epothilone analogue.
XX
SQ   Sequence 552 BP; 70 A; 176 C; 214 G; 92 T; 0 other;

  Query Match             46.1%;  Score 46.6;  DB 24;  Length 552;
  Best Local Similarity   66.3%;  Pred. No. 7.9e-05;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || | |   |||||  |  | || |||||   ||||||   | |||||||  ||||   |
Db     349 GCCTTGCACCTGCATGCGTTGACACGCGAAGCGCCGCTCTCCTTCTTCGTGCTGTACGCT 408

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           ||||   | ||  |||| ||||||   ||||||||||| |||
Db     409 TCGGGAGCAGGGCTCTTGGGCTCGCCGGGCCAGGGCAACTA 449


RESULT 6
AAF90037
ID   AAF90037 standard; DNA; 6462 BP.
XX
AC   AAF90037;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..6462
FT                   /*tag= a
FT                   /product= "type I polyketide synthase"
FT                   /transl_except= "(...3, aa: Met)"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83974.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 307-309; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 6462 BP; 1073 A; 2067 C; 2135 G; 1187 T; 0 other;

  Query Match             46.1%;  Score 46.6;  DB 22;  Length 6462;
  Best Local Similarity   66.3%;  Pred. No. 0.00011;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||   || ||||||||    ||||| || |  ||||| || || |||
Db    5623 GCCTGGAACCTGCACTTGCTCACCCGCGACTGCCCGCTCGATCATTTCGTGCTCTTCTCC 5682

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || ||  | ||  |  | ||| ||  || ||||||  | |||
Db    5683 TCCGCTGCAGGACTGCTGGGCGCGCCCGCCCAGGGAAACTA 5723


RESULT 7
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT   CDS             14056..14136
FT                   /*tag= b
FT   CDS             14148..15827
FT                   /*tag= c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
```

```
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             46.1%;  Score 46.6;  DB 18;  Length 15872;
  Best Local Similarity   66.3%;  Pred. No. 0.00013;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||   || |||    | | || ||| |||  |||||| |  || |||
Db    7682 GCGTGGAACCTGCACGAGCTCACCAAGCACCTGCGGCTCGACGCCTTCGTCCGCTTCTCC 7741

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || |  || ||| || ||||| |   |||||||| || |||
Db    7742 TCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTA 7782


RESULT 8
AAF90033
ID   AAF90033 standard; DNA; 34071 BP.
XX
AC   AAF90033;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Example 14; Page 289-300; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 coding strand), which encodes different
CC   polyketide synthases.
XX
SQ   Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

  Query Match             46.1%;  Score 46.6;  DB 22;  Length 34071;
  Best Local Similarity   66.3%;  Pred. No. 0.00014;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||   || |||||||||    ||||| || |  ||||| || || |||
Db   23846 GCCTGGAACCTGCACTTGCTCACCCGCGACTGCCCGCTCGATCATTTCGTGCTCTTCTCC 23905

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || ||  | ||  |  | ||| ||  || |||||| | |||
Db   23906 TCCGCTGCAGGACTGCTGGGCGCGCCCGCCCAGGGAAACTA 23946


RESULT 9
AAF90032/c
ID   AAF90032 standard; DNA; 42717 BP.
XX
AC   AAF90032;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Example 14; Page 274-288; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
```

```
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC   encodes different polyketide synthases.
XX
SQ   Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

  Query Match             46.1%;  Score 46.6;  DB 22;  Length 42717;
  Best Local Similarity   66.3%;  Pred. No. 0.00014;
  Matches   67;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || |||   ||||||   || ||||||||    ||||| || |  ||||| || || |||
Db   10274 GCCTGGAACCTGCACTTGCTCACCCGCGACTGCCCGCTCGATCATTTCGTGCTCTTCTCC 10215

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || ||  | ||  |  | ||| ||  || |||||| | |||
Db   10214 TCCGCTGCAGGACTGCTGGGCGCGCCCGCCCAGGGAAACTA 10174


RESULT 10
AAI68064
ID   AAI68064 standard; DNA; 552 BP.
XX
AC   AAI68064;
XX
DT   13-MAR-2002  (first entry)
XX
DE   9-keto-epothilone PKS inactive KR domain nucleotide sequence.
XX
KW   Cystobacterineae; recombinant; polyketide synthase; PKS; polyketide;
KW   epothilone; pilA gene; cytostatic; antipsoriatic; antiarthritic;
KW   antiarteriosclerotic; antiinflammatory; neuroprotective; vasotropic; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..552
FT                   /*tag= a
FT                   /note= "mutated KR domain of extender module 6"
XX
PN   WO200183800-A2.
XX
PD   08-NOV-2001.
XX
PF   26-APR-2001; 2001WO-US13793.
XX
PR   28-APR-2000; 2000US-0560367.
PR   14-SEP-2000; 2000US-232696P.
PR   21-DEC-2000; 2000US-257517P.
PR   03-APR-2001; 2001US-0825856.
PR   03-APR-2001; 2001US-0825876.
PR   13-APR-2001; 2001US-269020P.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Arslanian RL,  Ashley G,  Frykman S,  Julien B,  Katz L,  Khosla C;
PI   Lau J,  Licardi PJ,  Regentin R,  Santi D,  Tang L;
XX
DR   WPI; 2002-075167/10.
DR   P-PSDB; ABB07168.
XX
PT   Recombinant host cells useful for producing polyketides e.g. epothilone
PT   or its derivatives, comprises a recombinant expression vector encoding
PT   a heterologous polyketide synthase gene -
XX
PS   Example 11; Page 165-166; 221pp; English.
XX
CC   The invention provides a recombinant host cell, of the suborder
CC   Cystobacterineae comprising a recombinant expression vector encoding a
CC   heterologous polyketide synthase (PKS) gene and produces a polyketide
CC   synthesized by the PKS enzyme encoded on the vector. An epothilone
CC   derivative of a specified formula can be produced by culturing the host
CC   cell with a diketide equivalent compound of a specified formula. The host
CC   cells produces polyketides at high levels and are used in the production
CC   of not only epothilones, including new epothilone derivatives, but also
CC   other polyketides. Methods of purifying the epothilone derivatives are
CC   also useful for treating cancer, hyperproliferative diseases and
CC   conditions such as psoriasis, inflammation, sarcomas, neoplasms,
CC   lymphomas, multiple sclerosis, rheumatoid arthritis, atherosclerosis and
CC   /or restenosis. It improves polyketide production in any organism and
CC   also for production of products of recombinant PKS genes and modification
CC   enzymes. The present sequence represents the nucleotide sequence of a
CC   mutated and inactive KR domain of extender module 6 of the novel 9-keto-
CC   epothilone PKS of the present invention.
XX
SQ   Sequence 552 BP; 70 A; 177 C; 212 G; 93 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 24;  Length 552;
  Best Local Similarity   66.0%;  Pred. No. 0.00016;
  Matches   66;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           || | |   |||||  |  | || |||||   ||||||   | |||||| ||||   |
Db     349 GCCTTGCACCTGCATGCGTTGACACGCGAAGCGCCGCTCTCCTTCTTCGTGCTGTACGCT 408

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCT 100
           ||||   | ||  |||| ||||||   |||||||||| ||
Db     409 TCGGGAGCAGGGCTCTTGGGCTCGCCGGGCCAGGGCAACT 448


RESULT 11
AAV21186
ID   AAV21186 standard; DNA; 5676 BP.
XX
AC   AAV21186;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..5676
FT                   /*tag= a
FT                   /product= "polyketide synthase"
FT                   /note= "no stop codon given"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52844.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 38-43; 205pp; English.
XX
CC   The present sequence represents Amycolatopsis mediterranei strain wt3136
CC   5.7 kb KpnI fragment DNA, from the present invention. The present
```

```
CC   invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC   gene cluster DNA fragment comprising a DNA region involved directly or
CC   indirectly in the gene cluster responsible for rifamycin synthesis,
CC   including the adjacent DNA regions to the right and left which, by
CC   reason of their function in connection with rifamycin biosynthesis,
CC   qualify as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 19;  Length 5676;
  Best Local Similarity   68.5%;  Pred. No. 0.00022;
  Matches   63;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||  ||   || |||||||| | |   ||   |   |||||  |||| ||||| || ||
Db    2631 CTCGACGAGCTGACCCGCGAACTGGATCTTCGCGCGTTCGTCGTGTTCTCCTCCGCCTCC 2690

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || ||||||||||||   ||| ||||||| |||
Db    2691 GGGGTCTTCGGCTCCGCCGGGCAGGGCAACTA 2722


RESULT 12
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             45.1%;  Score 45.6;  DB 22;  Length 27541;
  Best Local Similarity   68.5%;  Pred. No. 0.00027;
  Matches   63;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||||||   || |||  ||| | |   || | |   ||||| ||||| |||||  || |
Db   24890 CTGCACGAGCTGACCGCCGACCTGGACCTCGCCGCTTTCGTGCTGTTCTCCTCCACGGCC 24949

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||||| ||||| |   ||| ||||||| |||
Db   24950 GGCGTCCTCGGCGCCGCCGGACAGGGCAACTA 24981


RESULT 13
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
```

```
FT                   /*tag= c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag= d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag= e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag= f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match            45.1%;  Score 45.6;  DB 19;  Length 53789;
  Best Local Similarity  68.5%;  Pred. No. 0.0003;
  Matches   63;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTCGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           ||  ||   || |||||||| | |   ||   |  |||||  |||| ||||| || ||
Db   45723 CTCGACGAGCTGACCCGCGAACTGGATCTTCGCGCGTTCGTCGTGTTCTCCTCCGCCTCC 45782

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           || ||||||||||   ||| ||||||| |||
Db   45783 GGGGTCTTCGGCTCCGCCGGGCAGGGCAACTA 45814


RESULT 14
AAF31641
ID   AAF31641 standard; DNA; 6333 BP.
XX
AC   AAF31641;
XX
DT   09-APR-2001  (first entry)
XX
DE   Mycobacterium tuberculosis mas gene.
XX
KW   Mycobacterium tuberculosis; attenuated microorganism;
KW   signature tagged transposon mutant; mutant library;
KW   mycobacterial infection; actinomycetales; antibacterial; immunostimulant;
KW   vaccine; mas; myocerosic acid synthase; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200102555-A1.
XX
PD   11-JAN-2001.
XX
PF   06-JUL-2000; 2000WO-IB00950.
XX
PR   06-JUL-1999;   99US-0142982.
PR   08-JUL-1999;   99US-0142833.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Gicquel B,  Guilhot C,  Camacho L;
XX
DR   WPI; 2001-091804/10.
DR   P-PSDB; AAB66471.
XX
PT   Screening a mutant library for mutants unable to grow under specific
PT   conditions and for identifying loci involved in pathogenicity,
PT   comprises using signature tagged transposon mutagenesis -
XX
PS   Example 8; Page 147-148; 159pp; English.
XX
CC   The present sequence is given in a specification relating to a method for
CC   screening a library of mutants. The method comprises constructing a
CC   library with insertions in genes and/or regulatory regions of the
CC   organisms of interest, where the insertion contains a tag and/or a
CC   transposon associated with a tag. The mutants are identified by
CC   hybridisation of the tags to known sequences. The method is useful for
CC   treating an individual suffering from a mycobacterial infection,
CC   suspected of being infected with a Mycobacterium, or having been
CC   exposed to an infectious Mycobacterium. It is also useful for
CC   identifying and isolating mutants of actinomycetales and for identifying
CC   compounds that have antibiotic activity. The method is used to identify
CC   mutants of microorganisms, preferably an actinomycetales, such as
CC   M. tuberculosis, M. bovis, M. leprae, M. avium, M. intracellulaire and
CC   M. paratuberculosis, that is unable to grow under specific conditions.
CC   It is especially useful for identifying loci involved in pathogenicity.
CC   It is useful in constructing vaccines. The method can be used to screen
CC   multiple libraries concurrently. It can screen libraries of different
CC   organisms or different strains of the same organism.
XX
SQ   Sequence 6333 BP; 1010 A; 2010 C; 2202 G; 1111 T; 0 other;

  Query Match            43.6%;  Score 44;  DB 22;  Length 6333;
  Best Local Similarity  65.0%;  Pred. No. 0.00067;
  Matches   65;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 CATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCT 61
           | |||   || |||  |  |||| ||  ||||||||||||   |||    |||| ||||
Db    5636 CCTGGAACCTACACCGCGCCACCCTCGGTCAGCCGCTGGACTGGTTCTGCTTGTTCTCCT 5695

Qy      62 CGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||   |||   |  ||||||||   || ||||||  |||
Db    5696 CGGGAGCGGCATTGCTCGGCTCGCCGGGTCAGGGCGCCTA 5735


RESULT 15
AAF90035
ID   AAF90035 standard; DNA; 8301 BP.
```

```
XX
AC   AAF90035;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..8301
FT                   /*tag=  a
FT                   /product= "type I polyketide synthase"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83972.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 302-305; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 8301 BP; 1426 A; 2629 C; 2669 G; 1577 T; 0 other;

  Query Match             43.6%;  Score 44;  DB 22;  Length 8301;
  Best Local Similarity   67.4%;  Pred. No. 0.0007;
  Matches   62;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      10 CTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCGGCGTCG 69
           || ||    |  |||||||| | |||||| ||| ||||||| || ||||| ||||| ||
Db    7549 CTCCATCATCACACCCGCGATCTGCCGCTCGACTTCTTCGTGCTCTTTTCATCGGCATCC 7608

Qy      70 GGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
             |||| ||  |   ||| ||||   ||||
Db    7609 TCGCTCTTAGGTCCTGCCGGGCAGGCAGGCTA 7640


Search completed: November  5, 2002, 12:35:35
Job time : 236.215 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_19950_20050
Perfect score: 101
Sequence: 1 gcatgggtgctgcacaccct.....tcgatcggccagggcagcta 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	68750	1	AF210843	AF210843 Sorangium
3		101	100.0	68750	6	AR193029	AR193029 Sequence
4		101	100.0	68750	6	AR199551	AR199551 Sequence
5		101	100.0	68750	6	AR199559	AR199559 Sequence
6		101	100.0	68750	6	AR199567	AR199567 Sequence
7		101	100.0	68750	6	AR201097	AR201097 Sequence
8		101	100.0	68750	6	AR208671	AR208671 Sequence
9		101	100.0	71989	6	AR172664	AR172664 Sequence
10	c	50	49.5	49736	1	AF319998	AF319998 Stigmatel
11		49.8	49.3	113193	1	AF357202	AF357202 Streptomy
12		48.2	47.7	16124	1	AX024384	AX024384 Sequence
13		48.2	47.7	16124	6	AX024277	AX024277 Sequence
14		48.2	47.7	123580	1	AF263912	AF263912 Streptomy
15		48.2	47.7	125401	6	AX211739	AX211739 Sequence
16	c	47.2	46.7	82746	1	AF453501	AF453501 Actinosyn
17		46.6	46.1	552	6	AX403009	AX403009 Sequence
18		46.6	46.1	6462	6	AX153794	AX153794 Sequence
19		46.6	46.1	34071	6	AX153790	AX153790 Sequence
20	c	46.6	46.1	42717	6	AX153789	AX153789 Sequence
21		45.6	45.1	552	6	AX403011	AX403011 Sequence
22		45.6	45.1	5676	6	A69718	A69718 Sequence 1
23	c	45.6	45.1	27522	1	AB070942	AB070942 Streptomy
24		45.6	45.1	27541	6	AX211706	AX211706 Sequence
25		45.6	45.1	53784	1	AMM223012	AJ223012 Amycolato
26		45.6	45.1	53789	6	A69720	A69720 Sequence 3
27	c	45.6	45.1	84985	1	SNA278573	AJ278573 Streptomy
28		45.6	45.1	90445	1	AF040570	AF040570 Amycolato
29		45	44.6	30000	6	AX250262	AX250262 Sequence
30		45	44.6	66808	1	SAU421825	AJ421825 Stigmatel
31		45	44.6	104326	1	AB070940	AB070940 Streptomy
32		44.6	44.2	10910	1	AX024383	AX024383 Sequence
33		44.6	44.2	10910	6	AX024276	AX024276 Sequence
34		44	43.6	6333	6	AX069081	AX069081 Sequence
35		44	43.6	8301	6	AX153792	AX153792 Sequence
36		44	43.6	9699	1	MSGMYACSYN	M95808 Mycobacteri
37	c	44	43.6	14326	1	AE007123	AE007123 Mycobacte
38	c	44	43.6	24292	1	MTCY24G1	Z83858 Mycobacteri
39		43.4	43.0	20394	1	SNA132222	AJ132222 Streptomy
40		43.4	43.0	20394	6	AX067996	AX067996 Sequence
41		43.4	43.0	30690	6	E38020	E38020 Avermectin
42		43.4	43.0	32870	1	AF007101	AF007101 Streptomy
43		43.4	43.0	64957	1	AB032367	AB032367 Streptomy
44		43.4	43.0	65140	6	AX211705	AX211705 Sequence
45		43.4	43.0	67523	1	SCU24241	U24241 Sorangium c

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:56"
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLFPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADRFIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAAAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
gene                 6260. .10492
                     /gene="epoB"
CDS                  6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene                 10639. .16137
                     /gene="epoC"
CDS                  10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILRLREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 5.5e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19950 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 20009

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db   20010 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 20050


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTLAYTDRE"
     CDS             6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVDTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
     gene            7610. .11875
                     /gene="epoA"
     CDS             7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRPIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFERLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
     misc_feature    7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
     misc_feature    9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
     misc_feature    10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
     misc_feature    11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
     gene            11872. .16104
                     /gene="epoP"
     CDS             11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
     gene            16251. .21749
                     /gene="epoB"
     CDS             16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match           100.0%;  Score 101;  DB 1;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           |||||||||||||||||||||||||||||||||||||||||||
Db   25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
```

```
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25562 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 25621

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  25622 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 25662


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match           100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 5.4e-16;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  19950 GCATGGGTGCTGCACACCCTTACCCGCGACCAGCCGCTGGACCTCTTCGTACTGTTTTCC 20009

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  20010 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 20050


RESULT 10
AF319998/c
LOCUS       AF319998               49736 bp    DNA     linear   BCT 28-MAY-2001
DEFINITION  Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF319998
VERSION     AF319998.1  GI:14210834
KEYWORDS    .
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Kunze,B., Blocker,H. and Muller,R.
  TITLE     Novel features in a combined polyketide synthase/non-ribosomal
            peptide synthetase: the myxalamid biosynthetic gene cluster of the
            myxobacterium Stigmatella aurantiaca Sga15
  JOURNAL   Chem. Biol. 8 (1), 59-69 (2001)
  MEDLINE   21110452
   PUBMED   11182319
REFERENCE   2  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Bloecker,H. and Mueller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
            38124, Germany
FEATURES             Location/Qualifiers
     source          1. .49736
                     /organism="Stigmatella aurantiaca"
                     /db_xref="taxon:41"
     CDS             complement(2. .1306)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aspartokinase"
                     /protein_id="AAK57182.1"
                     /db_xref="GI:14210835"
                     /translation="MRPIGSARQAGRGTSALALIVQKYGGTSVGDTERIKNVAKRCIA
                     AQKAGHDVVVVVSAMSGETNRLLKLVSQITERPNEREQDVVVATGEQVSIGLVALAIQ
                     TQKRKAVSFLGHQVRITTDSTFSKARIKSIDAERIVDALKKKNIVVVAGFQGQDEHGN
                     VTTLGRGGSDTTAVALAAALKADACEIYTDVDGVYTTDPNMVPAARKLERIAYEEMLE
                     LASVGAKVLQIRSVEFAMKYKVPLWVKSSFTDDPGTLVCEEDAAMEDVLVSGIAYDKN
                     EAKIAIRGVPDVPGVAAKIFGALDEQSIVVDLIVQNVSKDGRTDVTFTVGKADLTKAK
                     DVVKKIAKSVKAEGVETDDQVSKVSIVGVGMRNHSGVAAKMFTVLAGAGVNVQMISTS
                     EIKVSCVIHSNYTELAVRSLHTAFGLDKPEAGAVSEVVSLKG"
     CDS             1363. .2895
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative histidine ammonium lyase"
                     /protein_id="AAK57183.1"
                     /db_xref="GI:14210836"
                     /translation="MSRPRLNIDGDTLKLEEILQVARHTVTVELAPAAAARVKAARDL
                     VDRVAAGDTPSYGINTGFGTLAEVRIDKKDLRELQRNLILSHAAGVGSPLPLPEARVL
                     LLLRCNVLAKGYSGIRPETLALALEMLNRDVVPVVPERGSVGASGDLAPLAHLALVFI
                     GEGEAFYKGERLPAAQALERAGLKPVVLEAKEGLALVNGTQAMCAVGTLLQLRAEMLA
                     DLADLAGAMTLEGLLGSHKPFIPEIQDVRAHEGQKACAAHLRELLADSALVESHVNCS
                     KVQDPYSLRCMPQVHGAAREGLSFARRILEVEINSATDNPLVFVETERIVSGGNFHGQ
                     PVSLALDVAAMALTQLSAISERRVEQLVNPALSGLPPFLAKNSGLNSGFMIAQVTSAA
                     LVAESRVLSHPASVDSIPSSAGREDHVSMGMTAALKGRQVADFTRSCLAIELLVAAQA
                     LDYRQPTRAGKGPQAAYELIRSKIPTMEKDRELHRDIAAVSALIDSGELLNAVRTATR
                     GQ"
     gene            complement(3275. .7822)
                     /gene="mxaA"
     CDS             complement(3275. .7822)
                     /gene="mxaA"
                     /note="non-ribosomal peptide synthetase; contains a
                     condensing, adenylation, and PCP reductase domains"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaA"
                     /protein_id="AAK57184.1"
                     /db_xref="GI:14210837"
                     /translation="MSVRELLTTLAAKGVRLRLEGDKLQVEASKAVLTPELRAELSAN
                     KADILAFLRKHSSREETGSELGQPLARPEVIPLSSGQERLWFLDRLSPGTSQYNVHLG
                     LRARGALDTKALRRSLDELVRRHEVLRTHFPEVEGSPRQVIVSPDTGVALTIVELPGL
                     TESQREEEIYRRSDELAQQPFDLAKGPLLRVTVLALGPDDFALFVTKHHIITDGWSLG
```

```
                VFVRELSMLYASFVRGQAAALPPVSMQFADSALRERAWLAGESGARERAYWKGKLKGL
                PPLQLPVDHAVSTSTSHRGATVPLALSPALSDALRELASREGCSFFTVLFAAFSALLH
                RYSGQVDFGVGTVIANRGSVPADLIGFIANTLALRCDLSGEPTFSQWLARAHKVVLEA
                LDYQALPFSEVVQTVGASRDGGLNPLVRACFTLESIPAPTLDLPGTSWSFLNGAPDGS
                VEGVAKFELSLILAASEKGLAGMLEYSREVFDASTVERMVGHFQVLLESIVAHPEVPL
                SKLPLLTAEERGRLLNDWNGPVLDVPALCMHELVQAQAERTPQAVAVVSGQKTLTYAE
                LNRRANQLAHHLRRLGIQKEERVGLCVERTEDIVIGLLAILKAGGAYVPLDPAYPKER
                LALILEDAQVPVLLTQQRLVPELPATQARVVCLDKDWPTIGAESDTNPERITAPEAIA
                YLIYTSGSTGKPKGVMIEHRNAVAFLIWAMSVFSPKELAGTLASTSICFDLSVFEIFT
                PLCCGAKVIVAKNALELPELPAAREVTLINTVPSAMGALLRSGGVPSSVAIVNLAGEA
                LAGALVDQIYQLDHVRDVFNLYGPSETTTYSTFTRVNRGQTPTIGRPVGNTQVYALDS
                NREPMPIGVPGEVYIGGMGVARGYLGRPELTAERFVRSPFGGGSEARLYRTGDLARWL
                PDGQLEYLGRMDHQVKLRGFRIELGEIGAVLMEHSGIRDAVVVVREGLGADKQLVAYV
                VGRGEKAPEPAELRDYLKSKLPEYMVPFLFVGLDALPLTPNGKVDRAALPAPERTHSG
                PAKEHVAPRTPGEESLAAIWRQVLGVEQIGAHDNFFELGGHSLLLYRVLVLARSASGA
                DIPLRALLQAPTLEEMARAVEAAKTGSLPAHDVTVEMEADAVLDAEIALGKALPPVTG
                ALRTILLTGATGFLGAFLLEELCRRTDARIYCLVRSKTEQEGMNRIRKNLESYSLWNE
                ALAPRIVPVRGDIGQPLLGLSEKEFQRLSEEIDAIYHNGALVNFLYPYESMRAANVLG
                TREILRLATRTRIKPLHYVSTVSVLPLGRKAPIREDEPLEGPSSLVGGYAQSKWVAEK
                LVREASRRGLPVTILRPGRVTGHSRTGAWNTDDLVCRTLKGCVRMGVAPSVDALLDLT
                PVDYVSSAIVDLSMRPESIGQTYHLVNPQFVRADEMWNYMRAFGYGLRVLPYDQWLSE
                LGSAASSDSELGDLLMFLQQVPPEDRSVGGPRMVVCDSGDTLKALGGTGTSCPSVDAS
                LISTYLSSLVHRGFLKAPEVR"
gene            complement(7819. .11520)
                /gene="mxaB1"
CDS             complement(7819. .11520)
                /gene="mxaB1"
                /note="PKS; ketosynthase; acyltransferase"
                /codon_start=1
                /transl_table=11
                /product="MxaB1"
                /protein_id="AAK57185.1"
                /db_xref="GI:14210838"
                /translation="MRAGAGGHPLLGVSFTTSVQKENRFWESALSAKEPSYLADHRVG
                GAVVVPGAAYLEMALSAARELSGEKAYELVDTSFKEGLLLPEGQERVVQMALSGEGGD
                ALSFQLSSQKASTEPNAPAGWMAHVVGQIRPIEGGPSTQTPESFEAIRERCSEVIAQA
                AHYEELASRGVAYGPAFQGVQQIWRGRAEALGHVRLPEPLAARASAYRIHPALLDACF
                QLVAAAVPQEDLTPDAGPAVPVALAHLRLHEQPGTEVFCHVRLRPSEGARRGVYEFDL
                VLRDASGRILVEALGLRVQQLDAPAAEHAGKDLFFALEWRHASAVQQVPAEGPGASGR
                WLLIADGSELADTVESQLRERHGEVIRVDLAGAGQGRRVVDATSPQAFDGVLSEAFGA
                GSAPRGVIHLTGLEPREVDPIMGCGGVLHLVQALSRVSLASPPRLWLVTRGVHHGSGS
                RSASDVVQASLWGLGRTLAHEHSELKCTRVDVSAAMGTGDAGAALMRELLVADDEEEI
                SLRADGRHVGRIVRGSPGRAPSEAVVPAGGRPFHLERDAQGRLEFRAGARRPPGPGEL
                EIEVEVAALGGPGEGTGPERWEGAGEGIGCSGRVVAIGEGVKNVAVGEERIALVASGL
                GSHVSVPAGCTAIRPSSSSHGEAAALAATVLPAWYALFHLGRLEKRGRVWVLGATSGL
                GRAAIKLAKRAGAEVFAVATTDEQRASLRELGADQVMDPSDASFVSRVLDATGGRGVE
                LAIVAPGAVELERSLPTLSEGARVLDLRASGASARTGDANVAWCVVDVQDFARRRPER
                FASLWQEVKAAFEAGGLPSAAADAKALALKDPKARVVVPAADARRLRADGTYLVTGGL
                GGLGLAVAKWMVEQGAGHLVLVGRDTVLSPEQQEVVTALEAAGARVRVARADVSDRAQ
                IARILSEIAEGGAPLRGVIHAAGVLDDGVLLQQTVERFRRVMAPKVLGGWNLHVLTRE
                APLDFFVLYSSAASMFGAPGQGNYVAGNAFLDALAHHRRALGLPGLSINWGPFSEVGL
                AAAQANRGERLAQRGSDSLTPAEGNTILGRLLDGDVTQMAVMPLELRKWVEFYPRAKS
                SPWLSELVPADAGQEVGEARDVALLETLRTATPQEAKAALEQFVREQLARVLRLDSAR
                IDPEAPLQSFGLDSLMGLELRNRLVSGLGVSLPASLIWKHPTLAALCAHLLGEVMDRT
                LAETLAQAAAGAGPAVDESAADDKEVFVL"
gene            complement(11487. .14303)
                /gene="mxaB2"
CDS             complement(11487. .14303)
                /gene="mxaB2"
                /note="PKS; dehydratase; enoylreductase; ketoreductase;
                ACP"
                /codon_start=1
                /transl_table=11
                /product="MxaB2"
                /protein_id="AAK57186.1"
                /db_xref="GI:14210839"
                /translation="MSKQEKPAEQPSALQRAALLVEKMQARLDAVERARTEPIAVIGL
                GCRFPGGAVDGSSYWKILRDGVDALREVPESRWDVPGHFDSTRGVPGKMYGTRGGFLD
                DVEHFDAEFFGISPREAASLDPQQRLVLEVAWEALENAGLAPDQLAGSKTGVFMGVMS
                SDYMARLLKENDATRFDGYMATGNGYSFVPGRVSYVLGLQGPCMPVDTACSSSLVALH
                LASESLRAGESNLALAGGVNLILSPETMICLCSMQALASDGRCKTFDASADGYVRGEG
                CGVLVLKRLSDAQRDGDSILALIRGSAVNHDGASGGLTVPNGPSQQAVVQRALDNARI
                APALVGYIEAHGTGTPLGDPIELRALGAVLGKGRPADRPFFIGSVKTNIGHLEPAAGI
                AGVIKTILSLQHKEIPPHLHFRTPNPHVEWDRIPARVPVERVPWPAHEGRRIAGVSSF
                GLSGINAHVVLEEAPASTAPTAVAEEDKAQLVVLSAKTPRALSSLAEAWSSFLKNEEG
                GSLADISYTSALSRTPHEHRLALAAKSKQELIEQLEAFAAGKSHAGLSSGRRGTGLGQ
                KIVFVFPGQGSQWLGMGQRLYAENVTFREAIDWCHEAISRYTDWSLRDLLTNPGQQPR
                WNDIDVIQPTLFALQVALAAVWRWLGLEPHAVVGHSMGEVAAAHVAGALSLEDAARII
                CERSKLLRRVSGQGAMAVVDLSREQAEVEIRGAEDRIAIAVSNGPRSTVLSGDPAALK
                EVLERLERREVFCRWVKVDVASHSPQMDPLRQDLLDRMAAIRPTGSAVPLYSTVTGAP
                IDGRELSAPYWVRNLRQPVLFADAVQRLIQDGNVLFIELSPHPILVPFVDAMLREGDT
                SGRGLVVPSLIRDQEEWPSLLSSVGALFTRVIAWTGSGYIPTSVAGLRFLRTRGSASG
                TGWSCQQLRRRAGSCALAREVIPS"
gene            complement(14300. .29878)
                /gene="mxaC"
CDS             complement(14300. .29878)
                /gene="mxaC"
                /note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
                ACP regions"
                /codon_start=1
                /transl_table=11
                /product="MxaC"
                /protein_id="AAK57187.1"
                /db_xref="GI:14210840"
                /translation="MANRNDESNELSPAKRMLVALEKMQTRLNAVEGAAKEPIAIVGM
                ACRFPQADNVEAFWRLLQGGVDAVREIPLERWPKDEVADIGDGAMRWGGFLDSVDSFD
                PGFFGISPREAVRMDPQQRLLLEVAWEALEDAGLDVDKLSGSRSGVFIGACNDDYHCM
                QVERPETGDAFSATGVAASVLSGRLSYLFNLQGPSLVVDTACSSSLVSLHLACQSLRN
                RECNMALAGGVNLILSPQSVLLVSKLQALSPDGRSKAFDASANGFTRGEGCGIVVLKR
                LSDALADGDHILATIRGSAINQDGKSTGLTTPNVLSQQALIRQALESSGLKAEQVSYV
                EAHGTGTPLGDPIEAEALRETYGVPRPDGSVCGIGSVKTNVGHLESAAGIAGIMKVVL
                AMRHQSIPPHLHLKQVNPRIQLEGSALTIPTRLTPWKASGQPRRAGVSSFGISGTNAH
                VIIEEAPAPVAKDALPQRPEVLPLSARSREALVALAERYAERLGREGGGSLSDVCYTA
                SAHRTHHTHRLAVLADSREHTAERLRAFARGEQPSGAWSGRKAFGGRRKIVFVYPGQG
                AQRAGMGRALMESEPVFRDALVKVDKALQPHLGWSVLDEIAADEQRSRLAEIDVNQPV
                LFAVEVALTALWRSWGIEPDAVMGHSMGEVVAAHVAGALSLEDAVLIICHRSKLMRTL
                GGRGATMAMVELSVEEAEEAIRPWNGKIWLGGLNGPRSQVLSGEPEAVKEAVEALGKR
                GVFTRWVKMDVPSHTPLVEVISRELEGRIARITPRAADIPLCSTVTGEFLDGAKLDAA
                HWGNNLRNPVRFAAAAQHLLASDHGTFVEVSPHPVLLPAVEQTMSAGNTEGVVLASLR
                RNEDERHTMLSSLGALYSSGHRVAWNRVYPGKGQLVPLPTYPWQRERFWMEPAVRGVR
                ASRTSSGHPLIGERIDLSAQGEVSRLWQVQLGPKHFPLLAEHRVQGVPALPAAAMADM
                ALSAAKQTLGEPVSLEDVSFSALLSFPEDEERVLQLWLSTERPGASAFRLFSAPVMRE
                TPSDELTWALHAEGSLRHGDASVVEKAAPVALDELRARCGEEASVEAHYQTLKEIGLE
                YGPSFRVVRELYRAKGEALGRVQLEGELTAEEERYSLHPALLDGCFQVLAAALEGLDS
                EGGVYVPVALSRLRVLVRPEAALWSHVRIVSGAEAGSNTIECDVTLLNDQGRVVAEVR
                RLVCKKIAAASAPADEVDRWIYSAQWEPQQRQARAETGSSRTAGRWLVFADARGIGQS
                LASGLAERAGGCVVVEAGPSYERLEADRYRINPSKLEDFQALLKEAFSDQSPCKGMVH
                LYGLDVASNVTEPESFERALSLSSTSVVYLLQAWAQRGWRDTPRLWLVTTGVQVIASG
                EQPGSPAAAPLVGLAAVITHEHPELRCAHVDLSAQPSPEEIASLTEELLGDGAEDRIA
                LRGAVRHVARLGRGISGRKAAEPREFTADGTYLITGGLGGLGLEVAKWFIQRGARNLA
                LTGRGAPSEEAQKVLGELRANGAEVRVYRADVTSAADVARLLGSIDSELPPLRGVVHA
                AGVIADGVLTQLDEPRLRAVLAPKVQGAWNLHVQTQGRPLELFVMFSSAASMMGSPGQ
                GNYAAANAFMDVLAHHRRAAGLPGLSVNWGPWADVGLAAVATNRGDRLALRGVGNMPP
                AQALEALGRLLVQEAAQVAVVPLDLRQWREFYLSAAQSPFLSVLMQEKVSTPASRKGS
                ARELLAAAEPEQRKGILEGYLREQLGRILRMEPSRIEPDQPFGDLGMDSLTGLELRNR
                IESGLKLTLSATLIYAYPTVQALTTYLYGKLEPVLEAEARPTVRPTQTEPRVSVGVAP
```

Query Match 49.5%; Score 50; DB 1; Length 49736;
Best Local Similarity 69.4%; Pred. No. 0.0048;
Matches 68; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy       4 TGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCCTCG 63
           |||   |||||   ||| || ||||||  |||||| ||| ||||||| |||| ||| |||
Db    8631 TGGAACCTGCATGTCCTGACGCGCGAGGCGCCGCTCGACTTCTTCGTTCTGTATTCGTCG 8572

Qy      64 GCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
             ||| |     | |||||| |    || || |||| |||
Db    8571 GCGGCCTCGATGTTCGGCGCACCAGGACAAGGCAACTA 8534
```

```
RESULT 11
AF357202
LOCUS       AF357202              113193 bp    DNA     linear   BCT 17-JUL-2001
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
```

```
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRADEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene            complement(1805. .3628)
                     /gene="amphH"
     CDS             complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFERVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
     gene            3840. .4874
                     /gene="amphDIII"
     CDS             3840. .4874
                     /gene="amphDIII"
                     /note="mycosamine biosynthesis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphDIII"
                     /protein_id="AAK73500.1"
                     /db_xref="GI:14794892"
                     /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                     SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                     AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                     ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                     KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                     LDDLLLAHSW"
gene            5042. .33574
                /gene="amphI"
CDS             5042. .33574
                /gene="amphI"
                /note="polyketide synthase multienzyme polypeptide housing
                extension modules 9, 10, 11, 12, 13, and 14"
                /codon_start=1
                /transl_table=11
                /product="AmphI"
                /protein_id="AAK73501.1"
                /db_xref="GI:14794893"
                /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
                DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
                PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
                LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
                GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
                AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
                DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
                QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
                FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
                AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
                DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
                TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
                LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
                GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
                GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
                LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
                DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
                RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
                GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
                GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
                FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
                AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
                GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
                QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
                EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
                SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
                TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
                GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
                QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
                WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
                LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
                RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
                ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
                PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
                SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
                ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
                RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
                QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
                ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
                VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
                DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
                DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
                WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
                TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
```

```
TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTIWPETGRPRRAAVSS
FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH
```

```
Query Match             49.3%;  Score 49.8;  DB 1;  Length 113193;
Best Local Similarity   68.3%;  Pred. No. 0.005;
Matches   69;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || |||   ||||||  ||| |||||||| |     |||| |  |||||| || |  |||
Db  44342 GCCTGGCATCTGCACGACCTCACCCGCGATCTCGACCTGGCCGCCTTCGTCCTCTACTCC 44401

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          ||| | || ||||||||||||      || |||| || |||
Db  44402 TCGACCTCCGGCGTCTTCGGCAGCCCGGGTCAGGCCAACTA 44442


RESULT 12
AX024384
LOCUS       AX024384               16124 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024384
VERSION     AX024384.1  GI:10184588
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             47.7%;  Score 48.2;  DB 1;  Length 16124;
  Best Local Similarity   67.3%;  Pred. No. 0.015;
  Matches   68;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || | |    ||||||| || || |||||   ||| ||   | |||||||| ||||   |
Db   5430 GCCTTGCACTTGCACACGCTGACACGCGAAGCGCCTCTTTCCTTCTTCGTGCTGTACGCT 5489

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          || ||  | ||  | |||||||||   |||||||||| |||
Db   5490 TCTGCAGCTGGGCTTTTCGGCTCGCCAGGCCAGGGCAACTA 5530


RESULT 13
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match             47.7%;  Score 48.2;  DB 6;  Length 16124;
  Best Local Similarity   67.3%;  Pred. No. 0.015;
  Matches   68;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || | |    ||||||| || || |||||   ||| ||   | |||||||| ||||   |
Db   5430 GCCTTGCACTTGCACACGCTGACACGCGAAGCGCCTCTTTCCTTCTTCGTGCTGTACGCT 5489

Qy     61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
          || ||  | ||  | |||||||||   |||||||||| |||
Db   5490 TCTGCAGCTGGGCTTTTCGGCTCGCCAGGCCAGGGCAACTA 5530


RESULT 14
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
```

```
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .733)
                     /gene="nysF"
     CDS             complement(46. .733)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPQQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVHSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASFDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAFEKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFGRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                     QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                     LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                     DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                     SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
```

PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA

Query Match 47.7%; Score 48.2; DB 1; Length 123580;
Best Local Similarity 67.3%; Pred. No. 0.013;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || |||   ||||||  |   |||||||  |     |||| |  |||||| ||||  |||
Db   55337 GCCTGGCATCTGCACGACGCCACCCGCGGCCTCGACCTGGCCGGCTTCGTCCTGTACTCC 55396

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           ||||  || |||||| | |||     |||||||||| |||
Db   55397 TCGGTCTCCGGCGTCCTGGGCAGCCCCGGCCAGGGCAACTA 55437
```

RESULT 15
AX211739
LOCUS AX211739 125401 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 35 from Patent WO0159126.
ACCESSION AX211739
VERSION AX211739.1 GI:15523950
KEYWORDS .
SOURCE Streptomyces noursei.
ORGANISM Streptomyces noursei
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 125401)
AUTHORS Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
Gulliksen,O.M.
TITLE Gene cluster encoding a nystatin polyketide synthase and its
manipulation and utility
JOURNAL Patent: WO 0159126-A 35 16-AUG-2001;
Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
(NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
(NO)
FEATURES Location/Qualifiers
source 1. .125401
/organism="Streptomyces noursei"
/db_xref="taxon:1971"
/note="ATCC 11455"
BASE COUNT 15664 a 49692 c 42871 g 17174 t
ORIGIN

Query Match 47.7%; Score 48.2; DB 6; Length 125401;
Best Local Similarity 67.3%; Pred. No. 0.012;
Matches 68; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
Qy       1 GCATGGGTGCTGCACACCCTTACCCGCGAGCAGCCGCTGGACCTCTTCGTACTGTTTTCC 60
          || |||   ||||||  |   |||||||  |     |||| |  |||||| ||||  |||
Db   55745 GCCTGGCATCTGCACGACGCCACCCGCGGCCTCGACCTGGCCGGCTTCGTCCTGTACTCC 55804

Qy      61 TCGGCGTCGGGCGTCTTCGGCTCGATCGGCCAGGGCAGCTA 101
           ||||  || |||||| | |||     |||||||||| |||
Db   55805 TCGGTCTCCGGCGTCCTGGGCAGCCCCGGCCAGGGCAACTA 55845
```

Search completed: November 5, 2002, 12:11:44
Job time : 667.785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1113.44 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400
Perfect score: 101
Sequence: 1 ccggagcaggtggtgatcgc.....aagcggctgcatgtctccca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	40.2	39.8	512	10	AW163063	AW163063 au91g11.y
	2	38.8	38.4	1156	14	BQ922705	BQ922705 AGENCOURT
c	3	37.6	37.2	442	10	AW620463	AW620463 sj05e03.y
c	4	37.2	36.8	477	12	BF598970	BF598970 sv24b02.y
	5	36.8	36.4	512	9	AU183888	AU183888 AU183888
	6	36.6	36.2	1025	14	BQ926826	BQ926826 AGENCOURT
c	7	36.6	36.2	1065	14	BQ669726	BQ669726 AGENCOURT
	8	36.4	36.0	462	9	AU081551	AU081551 AU081551
c	9	36.4	36.0	1389	11	AY105590	AY105590 Zea mays
	10	36.2	35.8	1145	14	W29156	W29156 mb96a09.r1
c	11	36	35.6	420	10	BE588131	BE588131 WHE0657_G
	12	36	35.6	483	13	BJ254408	BJ254408 BJ254408
	13	35.6	35.2	554	17	AQ427466	AQ427466 CITBI-E1-
	14	35.6	35.2	596	13	BM656407	BM656407 170006873
	15	35.6	35.2	611	13	BM615342	BM615342 170006871
	16	35.6	35.2	613	13	BM605012	BM605012 170006870
	17	35.6	35.2	617	13	BM615807	BM615807 170006592
	18	35.6	35.2	632	13	BM601596	BM601596 170006870
	19	35.6	35.2	640	13	BM578146	BM578146 170006871
	20	35.6	35.2	641	13	BM584980	BM584980 170006872
	21	35.6	35.2	671	13	BM597907	BM597907 170006875
	22	35.6	35.2	688	13	BM596795	BM596795 170006874
	23	35.6	35.2	689	13	BM645847	BM645847 170006873
	24	35.6	35.2	694	13	BM597110	BM597110 170006874
c	25	35.6	35.2	764	17	AQ327293	AQ327293 nbxb0040P
	26	35.6	35.2	1255	14	BQ937459	BQ937459 AGENCOURT
c	27	35.4	35.0	385	13	BJ248257	BJ248257 BJ248257
c	28	35.4	35.0	694	11	AY108381	AY108381 Zea mays
c	29	35.4	35.0	831	12	BE889867	BE889867 601512119
	30	35.4	35.0	1660	17	AG058676	AG058676 Pan trogl
	31	35.2	34.9	535	13	BI597477	BI597477 603243696
	32	35.2	34.9	942	13	BI603297	BI603297 603245155
	33	35.2	34.9	1092	11	BC029543	BC029543 Homo sapi
	34	35.2	34.9	1235	13	BM547617	BM547617 AGENCOURT
	35	35	34.7	331	13	BJ471625	BJ471625 BJ471625
	36	35	34.7	545	13	BM488037	BM488037 pgm2n.pk0
c	37	35	34.7	621	9	AJ448803	AJ448803 AJ448803
	38	35	34.7	1114	17	AG082098	AG082098 Pan trogl
c	39	34.8	34.5	768	12	BG786329	BG786329 SEAUMC006
	40	34.6	34.3	686	17	BH576806	BH576806 BOGRK62TF
	41	34.6	34.3	925	17	CNS0091P	AL053013 Drosophil
c	42	34.6	34.3	1063	12	BF235933	BF235933 602025614
	43	34.4	34.1	612	10	BE418028	BE418028 SCL011.C1
c	44	34.4	34.1	1057	17	CNS046BO	AL276477 Tetraodon
	45	34.2	33.9	368	10	BE398539	BE398539 WHE0004.E

ALIGNMENTS

RESULT 1
AW163063
LOCUS AW163063 512 bp mRNA linear EST 09-NOV-1999
DEFINITION au91g11.y1 Schneider fetal brain 00004 Homo sapiens cDNA clone IMAGE:2783684 5' similar to SW:ID4_MOUSE P41139 DNA-BINDING PROTEIN INHIBITOR ID-4. [1] ;, mRNA sequence.
ACCESSION AW163063
VERSION AW163063.1 GI:6302096
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 512)
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: au91g11.x1
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

```
            Seq primer: -40RP from Gibco
            High quality sequence stop: 309.
FEATURES              Location/Qualifiers
     source           1. .512
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2783684"
                      /clone_lib="Schneider fetal brain 00004"
                      /sex="male"
                      /tissue_type="frontal lobe"
                      /dev_stage="5 months post-conception"
                      /lab_host="DH10B"
                      /note="Organ: brain; Vector: pBluescript SK (Stratagene);
                      Site_1: SstI; Site_2: XhoI; Double-stranded cDNA was
                      prepared from human fetal brain tissue.  5' and 3'
                      adaptors were used in cloning as follows: 5' adaptor
                      sequence:
                      5'-GAGAGAGAGAAGAGCTCAAGGATCCTTAATTAAATTAATCCCCCCCCCCCCC-3'
                      and 3' adaptor sequence:
                      5'-GAGAGAGAGACTCGAGTTTTTTTTTTTTTTTTT-3'.  The library was
                      size-selected for >0.5 kb inserts and has an average
                      insert size estimated at 1.2 kb.  This library was
                      constructed using the CAP-trapper method for full-length
                      enrichment and has not undergone amplification.    Library
                      was constructed by Dr. Claudio Schneider (LNCIB-Area
                      Science Park,  Trieste, Italy). "
BASE COUNT       77 a    143 c    192 g    100 t
ORIGIN

  Query Match             39.8%;  Score 40.2;  DB 10;  Length 512;
  Best Local Similarity   67.1%;  Pred. No. 10;
  Matches   57;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      13 GTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCGCGGC 72
           | | ||||  |||| ||||     || ||  |||||   |||||||||| ||| |||| ||
Db     339 GGGTTCGCTCGCGTAGAGCGCAGGGCGCGCGCGATGAAGGCGGGGTTCCCGGTGCGCCGC 398

Qy      73 GTGCGCACCAAGCGGCTGCATGTCT 97
           | ||  ||||  || | ||| ||
Db     399 TCGGGCCGCAAGGCGCCGTATGGCT 423


RESULT 2
BQ922705
LOCUS       BQ922705                1156 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8931561 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483876
            5', mRNA sequence.
ACCESSION   BQ922705
VERSION     BQ922705.1  GI:22337736
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTF
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2669  row: c  column: 13
            High quality sequence stop: 245.
FEATURES              Location/Qualifiers
     source           1. .1156
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:6483876"
                      /clone_lib="NIH_MGC_40"
                      /tissue_type="carcinoma, cell line"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                      Site_2: EcoRI; cDNA made by oligo-dT priming.
                      Directionally cloned into EcoRI/XhoI sites using the
                      following 5' adaptor: GGCACGAG(G). Library constructed  by
                      Ling Hong in the laboratory of Gerald M. Rubin (University
                      of California,  Berkeley) using ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life  Technologies).
                      Note: this is a NIH_MGC Library."
BASE COUNT      150 a    346 c    534 g    124 t      2 others
ORIGIN

  Query Match             38.4%;  Score 38.8;  DB 14;  Length 1156;
  Best Local Similarity   67.1%;  Pred. No. 19;
  Matches   55;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      19 GCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCGCGGCGTGCGC 78
           | |||   ||    |||||   ||||   ||||  ||||  |||||||||||||| ||||
Db     840 GGGGGACGGGGTGCAGCGGACGAGGCAGGCGCGAGGGGGGGCGCGGCGCGCGGCGAGCGC 899

Qy      79 ACCAAGCGGCTGCATGTCTCGC 100
           ||  ||||| | | | | |||
Db     900 GCCGGGCGGCGGAAGGGCGCGC 921


RESULT 3
AW620463/c
LOCUS       AW620463                442 bp    mRNA    linear   EST 03-DEC-2001
DEFINITION  sj05e03.y1 Gm-c1032 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1032-1013 5', mRNA sequence.
ACCESSION   AW620463
VERSION     AW620463.1  GI:7326665
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1099   Std Error: 0.00
            High quality sequence stop: 436.
FEATURES              Location/Qualifiers
     source           1. .442
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1032-1013"
                      /clone_lib="Gm-c1032"
                      /tissue_type="Cotyledons of 8-day-old 'Williams'
                      seedlings"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
```

XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 8-day-old 'Williams' seedlings which were propagated on paper towels with distilled water for 3 days (etiolated), then greenhouse grown for 5 days in potting soil. The cotyledons were flash-frozen in liquid nitrogen. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An 'anchor' nucleotide (V=A, C, or G) was added to the 3' end of the primer [GAGAGAGAGAGAGAGAGAACTAGTCTCGAG(T)18V] to anchor the primer at the 5' end of the poly(A) tract. After second strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI from Promega (40U/ul); all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then precipitated, redissolved, and ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 100% of the white and 87.5% of the blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=28 and 8 respectively). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

```
BASE COUNT       78 a    180 c     90 g     94 t
ORIGIN
```

```
  Query Match             37.2%;  Score 37.6;  DB 10;  Length 442;
  Best Local Similarity   61.0%;  Pred. No. 37;
  Matches   61;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          | |||| ||  ||||   ||||  | ||||  ||||| | |||||   ||||||| |
Db    282 CGGGAGAAGAGGGTGGAGGCGGCGGAGGAGGGAGCGGGGGAGGCGGCTGCGGCGGAGGAG 223

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
          | || |||  |   | | ||   | ||  ||  || ||||
Db    222 GAGGTGCGGTGGTCGAGGACTTCGAGGAAGCCGGTGTCGC 183
```

```
RESULT 4
BF598970/c
LOCUS       BF598970                 477 bp    mRNA    linear   EST 06-DEC-2001
DEFINITION  sv24b02.y1 Gm-c1057 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1057-436 5' similar to TR:Q9T078 Q9T078 HYPOTHETICAL 22.6 KD
            PROTEIN. ;, mRNA sequence.
ACCESSION   BF598970
VERSION     BF598970.1  GI:11691309
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact via email: ccu@resgen.com
            Insert Length: 1180   Std Error: 0.00
            High quality sequence stop: 421.
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1057-436"
                     /clone_lib="Gm-c1057"
                     /tissue_type="Degenerating cotyledons, 2 week old
                     seedling"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from mRNA isolated
                     from degenerating cotyledons of 2 week old seedlings from
                     PI468916. Complementary DNA was synthesized from mRNA
                     using a primer consisting of a poly(dT) sequence with a
                     XhoI restriction site. EcoRI adapters were ligated to the
                     blunt-ended cDNA fragments followed by XhoI digestion.
                     The cDNA fragments were directionally cloned into the
                     EcoRI-XhoI restriction site of the pBluescript vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (GibcoBRL). This library was constructed in the
                     laboratory of Dr. Randy Shoemaker."
BASE COUNT       88 a    191 c     95 g    103 t
ORIGIN
```

```
  Query Match             36.8%;  Score 37.2;  DB 12;  Length 477;
  Best Local Similarity   61.2%;  Pred. No. 45;
  Matches   60;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
          |||| ||  |||| . ||||  | ||||  ||||| | |||||   ||||||| |   |
Db    270 GGAGAAGAGGGTGGAGGCGGCGGAGGAGGGAGCGGGGGAGGCGGCTGCGGCGGAGGAGGA 211

Qy     63 GGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
          || |||  |   | | ||   | ||  ||  || ||||
Db    210 GGTGCGGTGGTCGAGGACTTCGAGGAAGCCGGTGTCGC 173
```

```
RESULT 5
AU183888
LOCUS       AU183888                 512 bp    mRNA    linear   EST 03-APR-2002
DEFINITION  AU183888 Rice cDNA from immature leaf including apical meristem
            (under short day condition) Oryza sativa (japonica cultivar-group)
            cDNA clone E60522, mRNA sequence.
ACCESSION   AU183888
VERSION     AU183888.1  GI:14191677
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from immature leaf including apical meristem (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
```

```
           PROJECT = 'RGP'.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="E60522"
                     /clone_lib="Rice cDNA from immature leaf including apical
                     meristem (under short day condition)"
                     /dev_stage="immature leaf including apical meristem (under
                     short day condition)"
BASE COUNT       78 a    150 c    176 g    107 t      1 others
ORIGIN

  Query Match             36.4%;  Score 36.8;  DB 9;  Length 512;
  Best Local Similarity   61.5%;  Pred. No. 55;
  Matches   59;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
           || |  || || |||  ||   | || |   |||| |  ||||||||||| ||| | | ||
Db     322 GGCGGGGGCGGCGATGCCGATGGCGGCGCTGGCGGGGGTGGCGATCGCGACGGCGCTGGC 381

Qy      63 GGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTC 98
           |||| | |   ||   ||   | || | || || ||
Db     382 GGCGGGGGTGCTGGTGACGGCGGGGTTCCACGTGTC 417


RESULT 6
BQ926826
LOCUS       BQ926826               1025 bp    mRNA    linear   EST 20-AUG-2002
DEFINITION  AGENCOURT_8773285 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6372361
            5', mRNA sequence.
ACCESSION   BQ926826
VERSION     BQ926826.1  GI:22341857
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1025)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2548  row: a  column: 02
            High quality sequence start: 18
            High quality sequence stop: 403.
FEATURES             Location/Qualifiers
     source          1. .1025
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6372361"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      146 a    273 c    498 g    107 t      1 others
ORIGIN

  Query Match             36.2%;  Score 36.6;  DB 14;  Length 1025;
  Best Local Similarity   65.1%;  Pred. No. 59;
  Matches   54;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
           |||| |||||| |   ||||| | || |   |||| |  ||||   | || ||||   ||
Db     683 GGAGTAGGTGGAGGGGGCGGGAGGGGGGGGGGCGGGGGGGCGGGGGGGGGGGGGGGGC 742

Qy      63 GGCGCGCGGCGTGCGCACCAAGC 85
           |||| || ||||||  | |  ||
Db     743 GGCGGGCCGCGTGCCGAGCGTGC 765


RESULT 7
BQ669726/c
LOCUS       BQ669726               1065 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8208383 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6254498
            5', mRNA sequence.
ACCESSION   BQ669726
VERSION     BQ669726.1  GI:21780560
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1065)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2404  row: b  column: 03
            High quality sequence stop: 221.
FEATURES             Location/Qualifiers
     source          1. .1065
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6254498"
                     /clone_lib="NIH_MGC_102"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT      109 a    551 c    254 g    151 t
ORIGIN

  Query Match             36.2%;  Score 36.6;  DB 14;  Length 1065;
  Best Local Similarity   62.6%;  Pred. No. 59;
  Matches   57;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           | |  | ||| || || || ||| |||| | ||| || |  ||||   |||| ||||| |
Db     623 CGGCGGGAGGGGGGGAGCGAGGGGGTGGGGTAAGGGGGGGGGGCGGGGGCGGGGGGGTCC 564

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
           | ||  || || | | |||    | ||  ||
Db     563 GGGGATCGGGGGGGGAGCAGGGGGGGGTCGC 533
```

```
RESULT 8
AU081551
LOCUS       AU081551                462 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AU081551 Rice panicle at flowering stage Oryza sativa (japonica
            cultivar-group) cDNA clone E0338_5A, mRNA sequence.
ACCESSION   AU081551
VERSION     AU081551.1  GI:6449609
KEYWORDS    EST.
SOURCE      Oryza sativa (japonica cultivar-group).
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Sasaki,T. and Yamamoto,K.
  TITLE     Rice cDNA from panicle at flowering stage
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="E0338_5A"
                     /clone_lib="Rice panicle at flowering stage"
                     /dev_stage="flowering stage"
                     /note="Organ: panicle; Rice cDNA from panicle at flowering
                     stage"
BASE COUNT       45 a    166 c    187 g     61 t      3 others
ORIGIN

  Query Match             36.0%;  Score 36.4;  DB 9;  Length 462;
  Best Local Similarity   66.7%;  Pred. No. 68;
  Matches   52;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      12 GGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGG 71
           |||||||| ||| | || ||   ||| || |||||  |||||| |  |   |||||||||
Db     140 GGTGATCGTGGGGGCGGGGCCGTCGGGGCTGGCGACGGCGGCGTGCCTGAAGGCGCGCGG 199

Qy      72 CGTGCGCACCAAGCGGCT 89
             ||||   |   |  |||
Db     200 GGTGCCGTCGCTGGTGCT 217


RESULT 9
AY105590/c
LOCUS       AY105590                1389 bp    mRNA    linear   HTC 25-MAY-2002
DEFINITION  Zea mays PCO144805 mRNA sequence.
ACCESSION   AY105590
VERSION     AY105590.1  GI:21208668
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1389)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1389)
  AUTHORS   Coe,E.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .1389
                     /organism="Zea mays"
                     /db_xref="MaizeDB:638562"
                     /db_xref="taxon:4577"
                     /clone="PCO144805"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      316 a    383 c    389 g    301 t
ORIGIN

  Query Match             36.0%;  Score 36.4;  DB 11;  Length 1389;
  Best Local Similarity   61.7%;  Pred. No. 64;
  Matches   58;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       6 GCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGC 65
           | ||| ||     | || || || |   |||| |  ||| || |||| |||| ||| ||
Db     189 GGAGGAGGAAGAGGAGGACGCGGCGGCGGCGGCGGCGGCCATGGCGGTGGGGATCGGGGA 130

Qy      66 GCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCG 99
           | | |||| ||| || | | | | |   || |||
Db     129 GGGAGGCGAGCGGACGAGGGGTCGGTCGGTGTCG 96


RESULT 10
W29156
LOCUS       W29156                  1145 bp    mRNA    linear   EST 11-SEP-1996
DEFINITION  mb96a09.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:337240 5', mRNA sequence.
ACCESSION   W29156
VERSION     W29156.1  GI:1309508
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1145)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:218640
            Seq primer: ETPrimer
            High quality sequence stop: 338.
FEATURES             Location/Qualifiers
     source          1. .1145
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:337240"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
```

was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)."
BASE COUNT 193 a 289 c 388 g 256 t 19 others
ORIGIN

Query Match 35.8%; Score 36.2; DB 14; Length 1145;
Best Local Similarity 61.5%; Pred. No. 72;
Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```
Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           ||||    || || |   | ||||| |      | || || ||||  || ||  ||||
Db    1015 CCGGGCGGGGCGGGGGCGGNGGGCGGGCGCGCGGGGGGGCGGGCGCGCGGGGGCGGGTGG 1074

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
           ||||||  ||| | |||| ||  ||||| ||
Db    1075 GCGGCGGNCGGGGGGCGCGCCGCGCGGCGGC 1105
```

RESULT 11
BE588131/c
LOCUS BE588131 420 bp mRNA linear EST 17-AUG-2000
DEFINITION WHE0657_G04_G04ZM Secale cereale root tip cDNA library Secale cereale cDNA clone WHE0657_G04_G04, mRNA sequence.
ACCESSION BE588131
VERSION BE588131.1 GI:9841163
KEYWORDS EST.
SOURCE rye.
ORGANISM Secale cereale
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Secale.
REFERENCE 1 (bases 1 to 420)
AUTHORS Anderson,O.D., Butler,E., Chao,S., Gustafson,J.P., Han,P.S., Hsia ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Root tip cDNA library from rye (Secale cereale)
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: M13 forward primer.
FEATURES Location/Qualifiers
source 1. .420
/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHE0657_G04_G04"
/clone_lib="Secale cereale root tip cDNA library"
/tissue_type="Root tip"
/dev_stage="Seedlings"
/lab_host="E. coli DH12S"
/note="Vector: pSPCRT2; Site_1: SalI; Site_2: NotI; Seeds were germinated aseptically on filter paper and transferred to a hydroponic growth system in a growth chamber when the primary root was 1 cm in length. Plants were grown in an environmental chamber until harvesting. The tissue, total RNA, and poly(A) RNA were prepared and a cDNA library was made (Butler and Gustafson) at University of Missouri, Columbia. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 48 a 221 c 85 g 66 t
ORIGIN

Query Match 35.6%; Score 36; DB 10; Length 420;
Best Local Similarity 70.6%; Pred. No. 84;
Matches 48; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```
Qy       6 GCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGC 65
           ||||||||||  | | || | || | |||||||| |||||   | |||||||| | |  |
Db     364 GCAGGTGGTGGCCCCAGGGGAGGCGGAAGCGGTGGAGGCGGCGGAGGCGGGGGTGGAAGA 305

Qy      66 GCGCGGCG 73
           ||||  ||
Db     304 GCGCCTCG 297
```

RESULT 12
BJ254408
LOCUS BJ254408 483 bp mRNA linear EST 08-APR-2002
DEFINITION BJ254408 Y. Ogihara unpublished cDNA library, Wh_f Triticum aestivum cDNA clone whf5i20 3', mRNA sequence.
ACCESSION BJ254408
VERSION BJ254408.1 GI:20078766
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 483)
AUTHORS Ogihara,Y. and Murai,K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .483
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whf5i20"
/clone_lib="Y. Ogihara unpublished cDNA library, Wh_f"
/tissue_type="spike at flowering date"
/dev_stage="Feekes' scale 10.5.1"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site_1: EcoRI; Site_2: XhoI; Plants were grown under hydroponic conditions at UC Davis, salt stressed for 12 hours, and for 7 days, then dissected and frozen (Akhunov in J Dvorak Lab). Total RNA was prepared from sheath tissue, equal quantities of RNA were pooled from the two samples, polyA was purified from the pooled RNA, a cDNA library was made, and the cDNA clones were in vivo excised to give pBluescript phagemids in the TJ Close lab at the University of California, Riverside (Akhunov, Chin , Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 91 a 161 c 129 g 98 t 4 others
ORIGIN

Query Match 35.6%; Score 36; DB 13; Length 483;
Best Local Similarity 63.5%; Pred. No. 83;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```
Qy      15 GATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCGCGGCGT 74
           || | |  ||||| |||   || ||||   || || | |  ||| |||| | ||| | ||
```

```
Db    319 GAGCTCCCGCGTGTAGCCGCCGCTGCACATGAACGTGCCACGGTACGCGCCCCGCAGGGT 378

Qy     75 GCGCACCAAGCGGCTGCATGTCTCG 99
          ||||| ||  |||||      ||||
Db    379 GCGCATCAGCCGGCTCTNCTCCTCG 403


RESULT 13
AQ427466
LOCUS       AQ427466                554 bp    DNA     linear   GSS 24-MAR-1999
DEFINITION  CITBI-E1-2555P11.TF CITBI-E1 Homo sapiens genomic clone 2555P11,
            DNA sequence.
ACCESSION   AQ427466
VERSION     AQ427466.1  GI:4500367
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 554)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2555P11.TR
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .554
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="2555P11"
                     /clone_lib="CITBI-E1"
                     /sex="male"
                     /cell_type="sperm"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
BASE COUNT       91 a    176 c    221 g     66 t
ORIGIN

  Query Match             35.2%;  Score 35.6;  DB 17;  Length 554;
  Best Local Similarity   64.6%;  Pred. No. 1e+02;
  Matches   53;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy      6 GCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGC 65
          ||||| ||||  |||| ||| ||| |  |||| || || |   || | || || |||||
Db    390 GCAGGAGGTGGGCGCGCGCGGGGACCCGGCGGGGCGGGGGCGGGCCGGGGAGTACGCGGT 449

Qy     66 GCGCGGCGTGCGCACCAAGCGG 87
          || ||  |  ||   || |||
Db    450 CCGGGGACTTAGCGGGAACCGG 471


RESULT 14
BM656407
LOCUS       BM656407                596 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687387858 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449645475 5', mRNA sequence.
ACCESSION   BM656407
VERSION     BM656407.1  GI:18955918
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 596)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU010049YY  row: J  column: 01
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .596
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449645475"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
BASE COUNT      125 a    175 c    175 g    121 t
ORIGIN

  Query Match             35.2%;  Score 35.6;  DB 13;  Length 596;
  Best Local Similarity   60.2%;  Pred. No. 1e+02;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      4 GAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCG 63
          |||| |||| |   | |   |  |  | ||| | || ||||| ||| |||   |||||
Db    445 GAGCGGGTGCTCCCCTCCATCACGACGGAAGTGCTGAAGGCGGTCGTGGCCCAGTTCGAT 504

Qy     64 GCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||| |||   ||  |||  ||||   |   || |||||
Db    505 GCGGGCGAGCTGATCACGCAGCGTGAGATGGTGTCGCA 542


RESULT 15
BM615342
LOCUS       BM615342                611 bp    mRNA    linear   EST 25-FEB-2002
DEFINITION  17000687146668 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            19600449698032 5', mRNA sequence.
ACCESSION   BM615342
VERSION     BM615342.1  GI:18913564
KEYWORDS    EST.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
            ,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
```

```
           Plate: NU01004AAO  row: G  column: 10
           Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Anopheles gambiae"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449698032"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT      127 a    180 c    180 g    124 t
ORIGIN

  Query Match             35.2%;  Score 35.6;  DB 13;  Length 611;
  Best Local Similarity   60.2%;  Pred. No. 1e+02;
  Matches   59;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       4 GAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCG 63
           |||| |||| |   | |   |  |  | ||| | || ||||| ||| |||   |||||
Db     465 GAGCGGGTGCTCCCCTCCATCACGACGGAAGTGCTGAAGGCGGTCGTGGCCCAGTTCGAT 524

Qy      64 GCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           ||| |||   ||  |||  ||||   |   || |||||
Db     525 GCGGGCGAGCTGATCACGCAGCGTGAGATGGTGTCGCA 562


Search completed: November  5, 2002, 16:17:34
Job time : 1119.44 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 81.2537 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400
Perfect score: 101
Sequence: 1 ccggagcaggtggtgatcgc.....aagcggctgcatgtctcgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	51.4	50.9	15872	10	US-09-861-289-1	Sequence 1, Appli
	2	32.4	32.1	925	10	US-09-823-038A-45	Sequence 45, Appl
c	3	32.4	32.1	1408	10	US-09-917-800A-1719	Sequence 1719, Ap
c	4	32.4	32.1	1739	10	US-09-789-836-10	Sequence 10, Appl
	5	32.2	31.9	379	10	US-09-867-550-233	Sequence 233, App
c	6	31.8	31.5	3205	10	US-09-969-347-173	Sequence 173, App
	7	31.8	31.5	7703	10	US-09-764-887-415	Sequence 415, App
	8	31.8	31.5	7703	10	US-09-764-887-456	Sequence 456, App
	9	31.6	31.3	2209	10	US-09-726-397A-2	Sequence 2, Appli
c	10	31.6	31.3	2366	10	US-09-726-397A-3	Sequence 3, Appli
c	11	31	30.7	2481	10	US-09-894-998-35	Sequence 35, Appl
	12	30.8	30.5	372	10	US-09-864-761-22031	Sequence 22031, A
c	13	30.8	30.5	1500	10	US-09-789-836-8	Sequence 8, Appli
c	14	30.6	30.3	1018	10	US-09-896-856-6	Sequence 6, Appli
c	15	30.6	30.3	1362	10	US-09-822-830A-19	Sequence 19, Appl
	16	30.6	30.3	1422	10	US-09-880-107-2208	Sequence 2208, Ap
c	17	30.6	30.3	1539	10	US-09-901-540-1	Sequence 1, Appli
	18	30.6	30.3	1539	10	US-09-901-540-2	Sequence 2, Appli
c	19	30.6	30.3	1539	10	US-09-901-257-1	Sequence 1, Appli
	20	30.6	30.3	1539	10	US-09-901-257-2	Sequence 2, Appli
c	21	30.4	30.1	423	10	US-09-833-790-52	Sequence 52, Appl
c	22	30.4	30.1	1231	10	US-09-880-107-3865	Sequence 3865, Ap
	23	30.4	30.1	1275	10	US-09-815-242-4042	Sequence 4042, Ap
c	24	30.4	30.1	1700	10	US-09-813-358-2	Sequence 2, Appli
	25	30.4	30.1	1785	10	US-09-732-357A-1	Sequence 1, Appli
	26	30.4	30.1	3715	10	US-09-880-107-2300	Sequence 2300, Ap
c	27	30.4	30.1	4223	12	US-10-001-215-1	Sequence 1, Appli
	28	30.4	30.1	4563	12	US-10-052-586-311	Sequence 311, App
	29	30.4	30.1	8894	10	US-09-764-864-1692	Sequence 1692, Ap
c	30	30.4	30.1	8894	10	US-09-764-877-3063	Sequence 3063, Ap
	31	30.4	30.1	15425	10	US-09-764-869-1654	Sequence 1654, Ap
	32	30.2	29.9	259	10	US-09-923-876-3423	Sequence 3423, Ap
	33	30.2	29.9	1797	10	US-09-969-708-328	Sequence 328, App
	34	30	29.7	426	10	US-09-960-352-12683	Sequence 12683, A
c	35	30	29.7	685	10	US-09-902-331-11	Sequence 11, Appl
c	36	30	29.7	776	10	US-09-902-331-3	Sequence 3, Appli
c	37	29.8	29.5	681	10	US-09-893-737-327	Sequence 327, App
c	38	29.6	29.3	1383	10	US-09-815-242-4002	Sequence 4002, Ap
	39	29.6	29.3	2072	10	US-09-772-105-5	Sequence 5, Appli
	40	29.6	29.3	2597	10	US-09-772-105-1	Sequence 1, Appli
	41	29.6	29.3	2834	10	US-09-954-456-286	Sequence 286, App
	42	29.6	29.3	2834	10	US-09-954-456-951	Sequence 951, App
	43	29.6	29.3	2834	10	US-09-954-456-1603	Sequence 1603, Ap
c	44	29.2	28.9	227	10	US-09-923-876-2979	Sequence 2979, Ap
	45	29.2	28.9	798	10	US-09-815-242-7941	Sequence 7941, Ap

ALIGNMENTS

RESULT 1
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-1

Query Match 50.9%; Score 51.4; DB 10; Length 15872;
Best Local Similarity 72.0%; Pred. No. 7.5e-05;
Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||||| | |||| |||| |||||||| | |  ||||||  | |||| |  ||
Db   10606 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 10665

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   ||||||| |||||| |  ||| ||||
Db   10666 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 10698
```

RESULT 2
US-09-823-038A-45
; Sequence 45, Application US/09823038A
; Patent No. US20020058335A1
; GENERAL INFORMATION:
; APPLICANT: Strachan, Lorna

```
;  APPLICANT:  Sleeman, Matthew
;  APPLICANT:  Abernethy, Nevin
;  APPLICANT:  Onrust, Rene
;  APPLICANT:  Kumble, Anand
;  APPLICANT:  Murison, Greg
;  TITLE OF INVENTION: Compositions Isolated From Stromal Cells
;  TITLE OF INVENTION:  and Methods For Their Use
;  FILE REFERENCE: 11000.1037c3
;  CURRENT APPLICATION NUMBER: US/09/823,038A
;  CURRENT FILING DATE:  2001-07-09
;  NUMBER OF SEQ ID NOS: 61
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
;   LENGTH: 925
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-823-038A-45

  Query Match             32.1%;  Score 32.4;  DB 10;  Length 925;
  Best Local Similarity   60.0%;  Pred. No. 3.1;
  Matches   54;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           ||   ||||   || |  ||||||||.|| |   || |||  |||   || ||||||    |
Db     395 CGCCCCAGGGCTTGGTGCCGGGCGTGGTGGCGGCCGTGGTGGCCGCCGTGGCGGGCGCGG 454

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
             | || ||  ||||  |    |||||| ||
Db     455 TGTCGAGCTTCGTGGCCTATCAGCGGCGGC 484


RESULT 3
US-09-917-800A-1719/c
; Sequence 1719, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1719
;   LENGTH: 1408
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM_024125
US-09-917-800A-1719

  Query Match             32.1%;  Score 32.4;  DB 10;  Length 1408;
  Best Local Similarity   60.0%;  Pred. No. 3;
  Matches   54;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           || |||||| || |  ||||| | ||| |   |||| ||  |   || |||  ||| | |
Db     650 CGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCGGGGCTCGGCGTCCCGGGCGGGCTGG 591

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
             |    || ||||  ||    || |||||
Db     590 ACGACGACGACGTGGACAGGCTGCCGCTGC 561


RESULT 4
US-09-789-836-10/c
; Sequence 10, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
;  APPLICANT: BRIGHAM, KENNETH L.
;  APPLICANT:  STECENKO, ARLENE A.
;  APPLICANT:  SEALY, LINDA
;  TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
;  FILE REFERENCE: N-6977
;  CURRENT APPLICATION NUMBER: US/09/789,836
;  CURRENT FILING DATE:  2001-02-20
;  PRIOR APPLICATION NUMBER: 60/183,584
;  PRIOR FILING DATE: 2000-02-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 1739
;   TYPE: DNA
;   ORGANISM: Rattus sp.
US-09-789-836-10

  Query Match             32.1%;  Score 32.4;  DB 10;  Length 1739;
  Best Local Similarity   60.0%;  Pred. No. 3;
  Matches   54;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           || |||||| || |  ||||| | ||| |   |||| ||  |   || |||  ||| | |
Db     736 CGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCGGGGCTCGGCGTCCCGGGCGGGCTGG 677

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
             |    || ||||  ||    || |||||
Db     676 ACGACGACGACGTGGACAGGCTGCCGCTGC 647


RESULT 5
US-09-867-550-233
; Sequence 233, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
;  APPLICANT: Leach, Martin D.
;  APPLICANT:  Mehraban, Fuad,
;  APPLICANT:  Conley, Pamela
;  APPLICANT:  Law, Debbie
;  APPLICANT:  Topper, James
;  TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells a
;  TITLE OF INVENTION:  Thereby
;  FILE REFERENCE: 21402-013 (Cura-313)
;  CURRENT APPLICATION NUMBER: US/09/867,550
;  CURRENT FILING DATE:  2001-09-20
;  PRIOR APPLICATION NUMBER: USSN 60/208,427
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 2125
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233
;   LENGTH: 379
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-867-550-233
```

```
  Query Match             31.9%;  Score 32.2;  DB 10;  Length 379;
  Best Local Similarity   61.2%;  Pred. No. 3.6;
  Matches   52;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          | |  || | ||  ||| || |||| || |   ||| ||     |||||||  |  ||||
Db    283 CAGCTGCTGCTGTGGATGGCCGGCGCGGTGGTCGCGCTGGGCTGGATCGCGATGCCGTTC 342

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGC 85
          |    || |||| ||||||  |  ||
Db    343 GTCCTGCTCGGCCTGCGCAGCCGGC 367


RESULT 6
US-09-969-347-173/c
; Sequence 173, Application US/09969347
; Patent No. US20020115085A1
; GENERAL INFORMATION:
;  APPLICANT: Ebner, Reinhard
;  TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-69
;  CURRENT APPLICATION NUMBER: US/09/969,347
;  CURRENT FILING DATE:  2001-10-02
;  PRIOR APPLICATION NUMBER: US/60/237,598
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: US/60/237,604
;  PRIOR FILING DATE: 2000-10-03
;  NUMBER OF SEQ ID NOS: 318
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 173
;   LENGTH: 3205
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-969-347-173

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 3205;
  Best Local Similarity   59.3%;  Pred. No. 4.1;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||  |   | || || || || || |||   | ||  | || |  || | ||||  |
Db    259 CCGGCGGCAGCGGCGACCGAGGACGAGGAGGCCGGGGCTCGGGAGTCCGGGCCGGGCCTG 200

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
          |||| |||  |  ||||    ||| ||  ||
Db    199 GCGGGGCGGAGGCTGCGGGGGAAGGGGAGGC 169


RESULT 7
US-09-764-887-415
; Sequence 415, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 415
;   LENGTH: 7703
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-415

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 7703;
  Best Local Similarity   64.0%;  Pred. No. 3.9;
  Matches   48;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
          |||||   | || |  |||||||| ||||  ||| |   ||| |  ||| |||||
Db   7529 CGGAGGCTGAGGCGGGCGCGGGCGGGGAGGCAGCAGCCGAGGAGGGCGCAGCGGGCCGAA 7588

Qy     62 CGGCGCGCGGCGTGC 76
          ||||| |||  ||
Db   7589 AGGCGCGGGGCCGGC 7603


RESULT 8
US-09-764-887-456
; Sequence 456, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA113
;  CURRENT APPLICATION NUMBER: US/09/764,887
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 658
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 456
;   LENGTH: 7703
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-887-456

  Query Match             31.5%;  Score 31.8;  DB 10;  Length 7703;
  Best Local Similarity   64.0%;  Pred. No. 3.9;
  Matches   48;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
          |||||   | || |  |||||||| ||||  ||| |   ||| |  ||| |||||
Db   7529 CGGAGGCTGAGGCGGGCGCGGGCGGGGAGGCAGCAGCCGAGGAGGGCGCAGCGGGCCGAA 7588

Qy     62 CGGCGCGCGGCGTGC 76
          ||||| |||  ||
Db   7589 AGGCGCGGGGCCGGC 7603


RESULT 9
US-09-726-397A-2
; Sequence 2, Application US/09726397A
; Patent No. US20020142011A1
; GENERAL INFORMATION:
;  APPLICANT: PARISH, TANYA
;  APPLICANT:  SMITH, DEBBIE
;  TITLE OF INVENTION: VACCINE
;  FILE REFERENCE: 117-329
;  CURRENT APPLICATION NUMBER: US/09/726,397A
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER: 0014845.20
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 2209
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (434)...(439)
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (452)...(704)
;   OTHER INFORMATION: unknown
US-09-726-397A-2

  Query Match             31.3%;  Score 31.6;  DB 10;  Length 2209;
  Best Local Similarity   58.5%;  Pred. No. 4.6;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;
```

```
Qy      3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
          ||||| | |||     |   | | ||||  |    | || | || |  || ||| | |
Db   1019 GGAGCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTGCCGGTGCCGCCGGTGCTGCG 1078

Qy     63 GGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTC 96
          || || ||||| | ||||||| |   | |  |||
Db   1079 GGTGCCCGGCGAGAGCACCAACCCCGTACTGGTC 1112


RESULT 10
US-09-726-397A-3/c
; Sequence 3, Application US/09726397A
; Patent No. US20020142011A1
; GENERAL INFORMATION:
;  APPLICANT: PARISH, TANYA
;  APPLICANT:  SMITH, DEBBIE
;  TITLE OF INVENTION: VACCINE
;  FILE REFERENCE: 117-329
;  CURRENT APPLICATION NUMBER: US/09/726,397A
;  CURRENT FILING DATE:  2000-12-01
;  PRIOR APPLICATION NUMBER: 0014845.20
;  PRIOR FILING DATE: 2000-06-16
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 2366
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1838)...(2090)
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (2103)...(2108)
;   OTHER INFORMATION: unknown
US-09-726-397A-3

  Query Match             31.3%;  Score 31.6;  DB 10;  Length 2366;
  Best Local Similarity   58.5%;  Pred. No. 4.6;
  Matches   55;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
          ||||| | |||     |   | | ||||  |    | || | || |  || ||| | |
Db   1523 GGAGCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTGCCGGTGCCGCCGGTGCTGCG 1464

Qy     63 GGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTC 96
          || || ||||| | ||||||| |   | |  |||
Db   1463 GGTGCCCGGCGAGAGCACCAACCCCGTACTGGTC 1430


RESULT 11
US-09-894-998-35/c
; Sequence 35, Application US/09894998
; Patent No. US20020090610A1
; GENERAL INFORMATION:
;  APPLICANT: Hosken, Nancy Ann
;  APPLICANT:  Craig H. Day
;  APPLICANT:  Davin C. Dillon
;  APPLICANT:  McGowan, Patrick
;  APPLICANT:  Sleath, Paul R.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION:  TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
;  FILE REFERENCE: 210121.538
;  CURRENT APPLICATION NUMBER: US/09/894,998
;  CURRENT FILING DATE:  2001-06-28
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
;   LENGTH: 2481
;   TYPE: DNA
;   ORGANISM: HSV-2
US-09-894-998-35

  Query Match             30.7%;  Score 31;  DB 10;  Length 2481;
  Best Local Similarity   64.8%;  Pred. No. 6.4;
  Matches   46;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
          || | ||| || |   |||| | |||| | |||| | ||| |   ||||||| |  |||
Db   1815 GGCGGAGGAGGAGGAGGCGGAGGAGGAGGAGGCGGAGGAGGAGGAGGCGGCGGCGACCGC 1756

Qy     63 GGCGCGCGGCG 73
          |||  | | ||
Db   1755 GGCCTGGGACG 1745


RESULT 12
US-09-864-761-22031
; Sequence 22031, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22031
;   LENGTH: 372
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC006385.3
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.9
```

```
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.6
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL =   6
;   OTHER INFORMATION: NT HIT: X87941.1, EVALUE 2.30e-01
;   OTHER INFORMATION: SWISSPROT HIT: P14404, EVALUE 2.70e+00
;   OTHER INFORMATION: EST_HUMAN HIT: AA286875.1, EVALUE 8.70e-02
US-09-864-761-22031

  Query Match             30.5%;  Score 30.8;  DB 10;  Length 372;
  Best Local Similarity   61.0%;  Pred. No. 7.8;
  Matches   50;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy       3 GGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGC 62
           |||| |||||| ||| | ||  |||||| | | | || ||| |   | || || | | |
Db     271 GGAGGAGGTGGAGATGGAGGAGGTGGAGGAGGTGATGGAGGAGGAGGTGGAGGAGCTGGA 330

Qy      63 GGCGCGCGGCGTGCGCACCAAG 84
           ||||  |    || | |  | |
Db     331 GGCGGCCCTTATGAGAAGAATG 352


RESULT 13
US-09-789-836-8/c
; Sequence 8, Application US/09789836
; Patent No. US20020082204A1
; GENERAL INFORMATION:
;  APPLICANT: BRIGHAM, KENNETH L.
;  APPLICANT:  STECENKO, ARLENE A.
;  APPLICANT:  SEALY, LINDA
;  TITLE OF INVENTION: TREATMENT OF INFLAMMATION WITH P20
;  FILE REFERENCE: N-6977
;  CURRENT APPLICATION NUMBER: US/09/789,836
;  CURRENT FILING DATE:  2001-02-20
;  PRIOR APPLICATION NUMBER: 60/183,584
;  PRIOR FILING DATE: 2000-02-18
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
;   LENGTH: 1500
;   TYPE: DNA
;   ORGANISM: Mus sp.
US-09-789-836-8

  Query Match             30.5%;  Score 30.8;  DB 10;  Length 1500;
  Best Local Similarity   58.9%;  Pred. No. 7.3;
  Matches   53;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           || |||||| || |  ||||| | ||| |   |||| ||  |   |  |||  ||| | |
Db     714 CGAAGCAGGCGGCGGGCGCGGCCTTGGCGTCGGCGGGGCTCGGCGTGCCGGGCGGGCTGG 655

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGC 91
             |    || ||||  ||    || |||||
Db     654 ACGACGACGACGTGGACAGGCTGCCGCTGC 625


RESULT 14
US-09-896-856-6/c
; Sequence 6, Application US/09896856
; Patent No. US20020137189A1
;   GENERAL INFORMATION:
;        APPLICANT: Baker, Joffre
;                   Chien, Kenneth
;                   King, Kathleen
;                   Pennica, Diane
;                   Wood, William
;        TITLE OF INVENTION: Cardiac Hypertrophy Factor and Uses Therefor
;        NUMBER OF SEQUENCES: 8
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 460 Point San Bruno Blvd
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/896,856
;             FILING DATE: 29-Jun-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US 08/733,850
;             FILING DATE: 18-OCT-1996
;             APPLICATION NUMBER: US 08/471,112
;             FILING DATE: 06-JUN-1995
;             APPLICATION NUMBER: 08/233,609
;             FILING DATE: 25-APR-1994
;             APPLICATION NUMBER: 08/286304
;             FILING DATE: 05-AUG-1994
;             APPLICATION NUMBER: 08/443129
;             FILING DATE: 17-MAY-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Conley, Deirdre L.
;             REGISTRATION NUMBER: 36,487
;             REFERENCE/DOCKET NUMBER: P0894P1D2C1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415/225-2066
;             TELEFAX: 415/952-9881
;             TELEX: 910/371-7168
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1018 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-896-856-6

  Query Match             30.3%;  Score 30.6;  DB 10;  Length 1018;
  Best Local Similarity   65.2%;  Pred. No. 8.3;
  Matches   45;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCG 68
           || || ||   |    |||| | | ||||   ||||| | |||||||||  | |||| ||
Db     542 GGCGGGGAAGACCCCGGTGGCGGAGGCGGCTGAGGCGGTGGCGGCGGGGGGCTCGGCCCG 483

Qy      69 CGGCGTGCG 77
            |||  |||
Db     482 GGGCCCGCG 474


RESULT 15
US-09-822-830A-19/c
; Sequence 19, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
;  APPLICANT: Genetics Institute, Inc.
;  APPLICANT:  Wong, Gordon G.
;  APPLICANT:  Clark, Hilary
;  APPLICANT:  Fechtel, Kim
;  APPLICANT:  Agostino, Michael J.
;  APPLICANT:  Howes, Steven H.
;  APPLICANT:  Resnick, Richard J.
;  APPLICANT:  Gulukota, Kamalakar
;  APPLICANT:  Graham, James R.
;  TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
;  FILE REFERENCE: GIN 6402
;  CURRENT APPLICATION NUMBER: US/09/822,830A
;  CURRENT FILING DATE:  2001-03-29
```

```
;  PRIOR APPLICATION NUMBER: 60/195,604
;  PRIOR FILING DATE: 2000-04-06
;  NUMBER OF SEQ ID NOS: 631
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 1362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-822-830A-19

  Query Match             30.3%;  Score 30.6;  DB 10;  Length 1362;
  Best Local Similarity   60.0%;  Pred. No. 8.2;
  Matches   51;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           ||| |  ||||||| | || |  |||| |   || |||  || |   | ||||| |   |
Db     607 CGGCGGCGGTGGTGGTGGCAGCGGTGGCGGTGGCAGTGGTGGTGGCGGTGGCGGTGGCAG 548

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCG 86
           | ||  ||  |  || | ||||| |
Db     547 CAGCCGGCCACCAGCCCCCCAAGAG 523


Search completed: November  5, 2002, 22:59:16
Job time : 89.2537 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 28.7371 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400
Perfect score: 101
Sequence: 1 ccggagcaggtggtgatcgc.....aagcggctgcatgtctcgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	101	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
	2	99.4	98.4	68750	3	US-09-335-409-1	Sequence 1, Appli
	3	99.4	98.4	68750	4	US-09-568-102-1	Sequence 1, Appli
	4	99.4	98.4	68750	4	US-09-567-969-1	Sequence 1, Appli
	5	99.4	98.4	68750	4	US-09-568-480-1	Sequence 1, Appli
	6	99.4	98.4	68750	4	US-09-568-486-1	Sequence 1, Appli
	7	99.4	98.4	68750	4	US-09-568-472-1	Sequence 1, Appli
	8	99.4	98.4	68750	4	US-09-567-899-1	Sequence 1, Appli
	9	51.4	50.9	15872	4	US-09-105-537-1	Sequence 1, Appli
	10	51.4	50.9	43280	2	US-08-804-227C-1	Sequence 1, Appli
	11	43.4	43.0	33529	4	US-09-144-085-3	Sequence 3, Appli
c	12	38.2	37.8	2721	6	5215881-2	Patent No. 5215881
c	13	38.2	37.8	8438	1	US-07-945-283-1	Sequence 1, Appli
	14	36.8	36.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	15	36.8	36.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	16	36.2	35.8	1505	1	US-07-915-246-1	Sequence 1, Appli
	17	35.4	35.0	80161	3	US-09-036-987A-1	Sequence 1, Appli
	18	35.4	35.0	80161	4	US-09-370-700-1	Sequence 1, Appli
	19	35.2	34.9	1088	3	US-09-040-285A-1	Sequence 1, Appli
	20	34.8	34.5	15393	4	US-09-453-702B-191	Sequence 191, App
	21	34.2	33.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
	22	34.2	33.9	44377	2	US-08-804-198-1	Sequence 1, Appli
	23	33	32.7	1030	3	US-08-858-003-2	Sequence 2, Appli
	24	33	32.7	1030	3	US-09-078-166-2	Sequence 2, Appli
	25	33	32.7	1030	4	US-08-997-467-2	Sequence 2, Appli
	26	32.8	32.5	7812	4	US-09-368-590-1	Sequence 1, Appli
c	27	31.8	31.5	3205	4	US-09-046-894-34	Sequence 34, Appl
	28	31.4	31.1	20235	1	US-07-642-734C-3	Sequence 3, Appli
	29	31.4	31.1	20235	3	US-08-439-009A-3	Sequence 3, Appli
c	30	31.2	30.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	31	31	30.7	1144	1	US-08-014-943A-1	Sequence 1, Appli
	32	31	30.7	1144	1	US-08-486-421-2	Sequence 2, Appli
	33	31	30.7	1144	1	US-08-470-911-2	Sequence 2, Appli
	34	31	30.7	1144	2	US-08-486-809-2	Sequence 2, Appli
c	35	31	30.7	1336	4	US-09-180-109A-28	Sequence 28, Appl
c	36	31	30.7	1420	4	US-09-180-109A-23	Sequence 23, Appl
	37	31	30.7	2214	3	US-08-864-038A-1	Sequence 1, Appli
	38	31	30.7	3331	3	US-08-864-038A-2	Sequence 2, Appli
	39	31	30.7	3331	3	US-08-864-038A-4	Sequence 4, Appli
c	40	30.8	30.5	1500	4	US-09-593-711A-10	Sequence 10, Appl
	41	30.8	30.5	1748	3	US-09-100-730-1	Sequence 1, Appli
	42	30.8	30.5	1926	4	US-09-249-585A-2	Sequence 2, Appli
	43	30.8	30.5	2580	3	US-09-050-863-2	Sequence 2, Appli
	44	30.8	30.5	2580	4	US-09-359-081-2	Sequence 2, Appli
c	45	30.8	30.5	5452	2	US-09-130-114-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
; APPLICANT: Kosan Biosciences, Inc.
; APPLICANT: Julien, Bryan
; APPLICANT: Katz, Leonard
; APPLICANT: Khosla, Chaitan
; APPLICANT: Tang, Li
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: Recombinant Methods and Materials for Producing
; TITLE OF INVENTION: Epothilone and Epothilone Derivatives
; FILE REFERENCE: 30062-20031.00
; CURRENT APPLICATION NUMBER: US/09/443,501A
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: US 60/130,560
; PRIOR FILING DATE: 1999-04-22
; PRIOR APPLICATION NUMBER: US 60/122,620
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: US 60/119,386
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: US 60/109,401
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 71989
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

Query Match 100.0%; Score 101; DB 4; Length 71989;
Best Local Similarity 100.0%; Pred. No. 2.2e-15;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18300 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 18359

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   18360 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 18400
```

RESULT 2
US-09-335-409-1
; Sequence 1, Application US/09335409

```
; Patent No. 6121029
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT: Ligon, James
;  APPLICANT: Molnar, Istvan
;  APPLICANT: Zirkle, Ross
;  APPLICANT: Cyr, Devon
;  APPLICANT: Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/335,409
;  CURRENT FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-335-409-1

  Query Match             98.4%;  Score 99.4;  DB 3;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 3
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,102
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 4
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 5
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012
```

```
RESULT 6
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 7
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 8
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             98.4%;  Score 99.4;  DB 4;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 5e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             50.9%;  Score 51.4;  DB 4;  Length 15872;
  Best Local Similarity   72.0%;  Pred. No. 0.00037;
  Matches   67;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
          |||||||||| | |||| |||| |||||||| | |  |||||| | |||| |  ||
Db  10606 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 10665

Qy     69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||  ||||||| |||||| |  ||| ||||
```

```
Db  10666 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 10698


RESULT 10
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             50.9%;  Score 51.4;  DB 2;  Length 43280;
  Best Local Similarity   72.0%;  Pred. No. 0.00036;
  Matches   67;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||||| | |||| |||| ||||||||   | |   ||||||  | |||| |  ||
Db   38468 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 38527

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   ||||||| |||||| |  ||| ||||
Db   38528 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 38560
```

```
RESULT 11
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             43.0%;  Score 43.4;  DB 4;  Length 33529;
  Best Local Similarity   64.4%;  Pred. No. 0.023;
  Matches   65;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           |||  ||||  |||| |    || |  ||    |||||||  ||| | |||  | || |
Db   22758 CCGATGCAGACGGTGCTGAGCGGGGACGAAGCGGCGGTGCTCGCGGTGGCGCGGCGGCTG 22817

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           | |||||  |||  |||||| |||| |||| ||| |||||
Db   22818 GAGGCGCAGGGCCGGCGCACGCGGCGTCTGCGTGTGTCGCA 22858


RESULT 12
5215881-2/c
;Patent No. 5215881
;    APPLICANT: CHEUNG, ANDREW K.
;    TITLE OF INVENTION: PSEUDORABIES DIAGNOSIS PROBES
;    NUMBER OF SEQUENCES: 3
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/537,855
;      FILING DATE: 13-JUN-1990
;      FILING DATE:
;SEQ ID NO:2:
;      LENGTH: 2721
5215881-2

  Query Match             37.8%;  Score 38.2;  DB 6;  Length 2721;
  Best Local Similarity   61.6%;  Pred. No. 0.39;
  Matches   61;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           |||  |  || || | | ||||  |||| | |||||| |  ||||  |||||||| |  |
Db    787 CGGCGGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGGGCT 728

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
           |||||  ||  |  ||  ||   |  || |   || |||
Db    727 CGGCGGCCGAGGATCGTCCCGGTCCCCTTCTCCTCCCGC 689


RESULT 13
US-07-945-283-1/c
; Sequence 1, Application US/07945283
; Patent No. 5352596
;  GENERAL INFORMATION:
;    APPLICANT:  Cheung, Andrew K.
;    APPLICANT:  Wesley, Ronald D.
;    TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
```

```
;    TITLE OF INVENTION:  Involving The EP0 and LLT Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis P. Ribando
;      STREET:  1815 No. 5352596th University Street
;      CITY:  Peoria
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  61604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,283
;      FILING DATE:  19920911
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ribando, Curtis P
;      REGISTRATION NUMBER:  27976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  309-685-4011 ext.513
;      TELEFAX:  309-685-4128
;   INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8438 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Pseudorabies virus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  622..6495
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1099, "g")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1267, "t")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1381, "c")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(1566, "c")
;    FEATURE:
;      NAME/KEY:  variation
;      LOCATION:  replace(7010, "g")
US-07-945-283-1

  Query Match             37.8%;  Score 38.2;  DB 1;  Length 8438;
  Best Local Similarity   61.6%;  Pred. No. 0.37;
  Matches   61;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy      2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
          |||  |  || || | | ||||  |||| | ||||||  |  ||||  ||||||||  |  |
Db   2130 CGGCGGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGGGCT 2071

Qy     62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
          |||||  ||  |  ||  ||   |  || |   || |||
Db   2070 CGGCGGCCGAGGATCGTCCCGGTCCCCTTCTCCTCCCGC 2032


RESULT 14
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             36.4%;  Score 36.8;  DB 4;  Length 4403765;
  Best Local Similarity   63.6%;  Pred. No. 0.56;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
          ||  |  |||| ||  |||||||  || || | |  |  ||| ||||||||||  || |
Db 835263 CGACGGCGGTGCTGGCCGCGGGCGCCGATCAGGTGTCGGTGGCCATCGCGGCGGCTTTTG 835322

Qy     62 CGGCGCGCGGCGTGCGCACCAAGCGGCT 89
          ||||| ||||  |  |    ||  |||
Db 835323 GGGCGCACGGCCAGGCCTATCAGGCGCT 835350


RESULT 15
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             36.4%;  Score 36.8;  DB 4;  Length 4411529;
  Best Local Similarity   63.6%;  Pred. No. 0.56;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
          ||  |  |||| ||  |||||||  || || | |  |  ||| ||||||||||  || |
Db 833085 CGACGGCGGTGCTGGCCGCGGGCGCCGATCAGGTGTCGGTGGCCATCGCGGCGGCTTTTG 833144

Qy     62 CGGCGCGCGGCGTGCGCACCAAGCGGCT 89
          ||||| ||||  |  |    ||  |||
Db 833145 GGGCGCACGGCCAGGCCTATCAGGCGCT 833172


Search completed: November  5, 2002, 13:38:04
Job time : 2032.74 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 137.215 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400
Perfect score: 101
Sequence: 1 ccggagcaggtggtgatcgc.....aagcggctgcatgtctcgca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	101	100.0	71989	21	AAA29349	Sorangium cellulos
2	99.4	98.4	68750	21	AAZ55887	Sorangium cellulos
3	51.4	50.9	15872	18	AAT68715	Streptomyces venez
4	51.4	50.9	15872	21	AAZ87283	S. venezuelae vep
5	51.4	50.9	20394	22	AAF24892	Pimaricin biosynth
6	51.4	50.9	43280	18	AAT80413	Tylactone synthase
7	43.4	43.0	33529	23	AAS17367	DNA sequence of S.
8	43.4	43.0	65140	22	AAD17184	Streptomyces nours
9	43.4	43.0	125401	22	AAD17186	Streptomyces nours
10	42.2	41.8	27541	22	AAD17185	Streptomyces nours
11	41.8	41.4	4851	22	AAH52062	Mycobacterium tube
12	40.2	39.8	31422	22	AAH79278	Streptomyces averm
13	38.6	38.2	31422	21	AAA92302	S. avermitilis ave
c 14	38.2	37.8	2721	12	AAQ10212	BamHI J-I fragment
c 15	38.2	37.8	2721	12	AAQ10543	BamHI J-I fragment
c 16	38.2	37.8	8438	15	AAQ73500	DNA encoding Pseud
17	37.6	37.2	77536	21	AAA14651	Nucleotide sequenc
18	37.2	36.8	696	24	ABQ33710	Oligonucleotide fo
c 19	37.2	36.8	696	24	ABQ33711	Oligonucleotide fo
20	37	36.6	3978	21	AAC55785	Type I polyketide
21	37	36.6	12249	21	AAC55840	Complete Mitomycin
22	37	36.6	18331	21	AAC55857	Complete nucleotid
23	36.8	36.4	4403765	22	AAI99683	Mycobacterium tube
24	36.8	36.4	4411529	22	AAI99682	Mycobacterium tube
25	36.2	35.8	1505	15	AAQ55750	Genomic clone G11F
c 26	36	35.6	88421	24	AAL40781	88421nt genomic DN
27	35.6	35.2	2685	24	ABK83742	Human cDNA differe
c 28	35.6	35.2	9692	22	AAS27739	DNA encoding novel
29	35.6	35.2	9730	22	AAS27740	DNA encoding novel
30	35.4	35.0	881	24	ABQ29954	Oligonucleotide fo
c 31	35.4	35.0	881	24	ABQ29955	Oligonucleotide fo
32	35.4	35.0	4209	22	AAH51966	Mycobacterium tube
33	35.4	35.0	4209	22	AAF31625	Mycobacterium tube
34	35.4	35.0	4260	22	AAF31645	Mycobacterium tube
35	35.4	35.0	14775	22	AAF88338	S. spinosa DNA fra
36	35.4	35.0	50000	22	AAF88313	S. spinosa DNA fra
37	35.4	35.0	50000	22	AAF88316	S. spinosa DNA fra
38	35.4	35.0	80161	20	AAZ21501	DNA fragment of Sa
39	35.2	34.9	1088	20	AAX08638	Sequence encoding
40	35.2	34.9	222930	24	ABK84349	Human cDNA differe
41	34.8	34.5	1167	22	ABA89147	Escherichia coli p
c 42	34.8	34.5	2349	22	AAH41951	Putative human CLA
c 43	34.8	34.5	48254	22	ABA89141	Escherichia coli p
c 44	34.8	34.5	48345	22	ABA89142	Escherichia coli p
c 45	34.6	34.3	2373	23	AAS65199	DNA encoding novel

ALIGNMENTS

RESULT 1
AAA29349
ID AAA29349 standard; DNA; 71989 BP.
XX
AC AAA29349;
XX
DT 12-SEP-2000 (first entry)
XX
DE Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 3..992
FT /*tag= a
FT /label= ORF_A
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 989..1501
FT /*tag= b
FT /label= ORF_B
FT /product= transposase
FT /note= "not part of the PKS"
FT CDS 1998..6263
FT /*tag= c
FT /label= epoA_gene
FT /note= "encodes the loading domain"
FT misc_RNA 2031..3548

```
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                         domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                         domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                         domain, potentially involved in formation of the
FT                         thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                         domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                         module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                         module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                         NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                         module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                         module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                         module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                         module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
```

```
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 101;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-15;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  18300 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 18359

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||||||||||||||||||||||||||||||
Db  18360 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 18400


RESULT 2
AAZ55887
ID   AAZ55887 standard; DNA; 68750 BP.
XX
AC   AAZ55887;
XX
DT   10-APR-2000  (first entry)
XX
DE   Sorangium cellulosum 68.75 kb contig.
XX
KW   Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW   anticancer; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1826
FT                   /*tag= a
FT                   /partial
FT                   /product= "Partial Orf 1 protein (AAY58580)"
FT                   /note= "No initiation codon given in the specification"
FT   CDS             complement (1900..3171)
FT                   /*tag= b
FT                   /product= "Orf 2 protein (AAY58581)"
FT   CDS             3415..5556
FT                   /*tag= c
FT                   /product= "Orf 3 protein (AAY58582)"
FT   CDS             complement (5612..5992)
FT                   /*tag= d
FT                   /product= "Orf 4 protein (AAY58583)"
FT   CDS             6226..6675
FT                   /*tag= e
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853..64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
```

```
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR          AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR          AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR          AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             98.4%;  Score 99.4;  DB 21;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 9.6e-15;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTC 23971

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||||||||||||| |||||||||||||||||||||||||||
Db   23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 3
AAT68715
ID   AAT68715 standard; DNA; 15872 BP.
XX
AC   AAT68715;
XX
DT   01-SEP-1997  (first entry)
XX
DE   Streptomyces venezuelae polyketide synthase vep ORF1.
XX
KW   Polyketide synthase; polyhydroxyalkanoate monomer synthase;
KW   polyhydroxybutyrate; biodegradable polymer; vep gene;
KW   metabolic engineering; ss.
XX
OS   Streptomyces venezuelae.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
FT                   /*tag= a
FT   CDS             14056..14136
FT                   /*tag= b
FT   CDS             14148..15827
FT                   /*tag= c
XX
PN   WO9722711-A1.
XX
PD   26-JUN-1997.
XX
PF   18-DEC-1996;   96WO-US20119.
XX
PR   19-DEC-1995;   95US-0008847.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Williams MD,  Xue Y;
XX
DR   WPI; 1997-341701/31.
DR   P-PSDB; AAW19629-30 AND AAW00918.
XX
PT   Expression cassettes for production of polyhydroxyalkanoate(s) -
PT   provide wide range of biodegradable polymers for medical or
PT   industrial use
XX
PS   Claim 54; Fig 23; 91pp; English.
XX
CC   Streptomyces venezuelae vep ORF1 (AAT68715) comprises the polyketide
CC   synthase (PKS) gene cluster encoding a polyene of 12 carbons (see
CC   also AAW19629-30 and AAW00918).  It contains 5 PKS modules, with a 5'
CC   loading module and a 3' end domain.  Each of the sequenced modules
CC   includes a keto-ACP, an acyltransferase, a dehydratase, a keto-
CC   reductase and an acyl carrier protein domain.  The gene cluster was
CC   cloned using a heterologous hybridisation strategy from a genomic
CC   DNA library.  A novel expression cassette encoding the first module
CC   from the vep gene cluster and module 7 from the Streptomyces tylP
CC   gene cluster has polyhydroxyalkanoate (PHA) monomer synthase
CC   activity and can be used for PHA prodn. in host (esp. insect) cells
CC   for use as a biodegradable polymer.
XX
SQ   Sequence 15872 BP; 2093 A; 6313 C; 5503 G; 1963 T; 0 other;

  Query Match             50.9%;  Score 51.4;  DB 18;  Length 15872;
  Best Local Similarity   72.0%;  Pred. No. 0.0014;
  Matches   67;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||||| | |||| |||| ||||||||   | |   ||||||  | |||| |  ||
Db   10606 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 10665

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||   ||||||| |||||| |  ||| ||||
Db   10666 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 10698


RESULT 4
AAZ87283
ID   AAZ87283 standard; DNA; 15872 BP.
XX
AC   AAZ87283;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae vep ORF 1, SEQ ID NO:1.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             20..13912
```

```
FT                   /*tag=  a
FT                   /product= "vep ORF 1 amino acid sequence #1 (AAY77177)"
FT   CDS             14056..14151
FT                   /*tag=  b
FT                   /product= "vep ORF 1 amino acid sequence #3 (AAY77199)"
FT   CDS             14167..15827
FT                   /*tag=  c
FT                   /product= "vep ORF 1 amino acid sequence #2 (AAY77178)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77177, AAY77178, AAY77199.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Example 3; Figure 23; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants. The
CC   present sequence represents a Streptomyces venezuelae ATCC 15439 DNA
CC   sequence, designated vep ORF 1 in the specification, which actually
CC   contains 3 open reading frames, which encode proteins AAY77177-Y77178 and
CC   AAY77199. The vep ORF 1 protein is defined in the specification as a PHA
CC   monomer synthase.
XX
SQ   Sequence 15872 BP; 2088 A; 6304 C; 5513 G; 1967 T; 0 other;

  Query Match             50.9%;  Score 51.4;  DB 21;  Length 15872;
  Best Local Similarity   72.0%;  Pred. No. 0.0014;
  Matches   67;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGCCGCG 68
           |||||||||| | |||| |||| ||||||||   | |   ||||||  | |||| |  ||
Db   10606 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCCGACG 10665

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   ||||||| |||||| |  ||| ||||
Db   10666 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 10698
```

```
RESULT 5
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag=  a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             50.9%;  Score 51.4;  DB 22;  Length 20394;
  Best Local Similarity   69.3%;  Pred. No. 0.0014;
  Matches   70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           |||  |  ||| |||||| ||||||  ||| | ||||||| ||| |||||||||   |||
Db   11944 CCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACGCGGTGCTGGCCATCGCGGCGCACTTC 12003

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||| |   ||       ||||  || ||||  ||||||||
Db   12004 GCGGGGGAGGGGCGCAAAACCACCCGACTGCGGGTCTCGCA 12044


RESULT 6
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
```

```
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             50.9%;  Score 51.4;  DB 18;  Length 43280;
  Best Local Similarity   72.0%;  Pred. No. 0.0013;
  Matches   67;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||||| | |||| |||| ||||||||   | |   ||||||  | |||| |  ||
Db   38468 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 38527

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   ||||||| |||||| |  ||| ||||
Db   38528 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 38560
```

```
RESULT 7
AAS17367
ID   AAS17367 standard; DNA; 33529 BP.
XX
AC   AAS17367;
XX
DT   12-MAR-2002  (first entry)
XX
DE   DNA sequence of S. cellulosum polyketide synthase cosmid, pKOS28-26.
XX
KW   Polyketide synthase; PKS; catalytic domain; ketosynthase domain;
KW   acyl transferase domain; dehydratase domain; ketoreductase domain;
KW   acyl carrier protein domain; pKOS28-26; ds.
XX
OS   Sorangium cellulosum.
XX
PN   US6280999-B1.
XX
PD   28-AUG-2001.
XX
PF   31-AUG-1998;   98US-0144085.
XX
PR   22-JAN-1998;   98US-0010809.
XX
PA   (KOSA-) KOSAN BIOSCIENCE.
XX
PI   Gustafsson C,  Betlach MC,  Ashley G,  Julien B,  Ziermann R;
XX
DR   WPI; 2001-606536/69.
XX
PT   Novel purified, isolated DNA molecule from Sorangium cellulosum having
PT   polyketide open reading frame encoding modules with one or more domains
PT   such as ketosynthase, acyl transferase and acyl carrier protein domains
PT   -
XX
PS   Claim 4; Fig 1; 72pp; English.
XX
CC   The present invention relates to the isolation of novel Sorangium
CC   cellulosum polyketide synthases (PKS), and the polynucleotide sequences
CC   encoding them. The polyketide synthases include catalytic domains such
CC   as ketosynthase domain, acyl transferase domain, dehydratase domain,
CC   ketoreductase domain and acyl carrier protein domain. A host cell
CC   comprising a PKS ORF (open reading frame) which encodes one or more
CC   more PKS domains is useful for producing polyketide synthases from which
CC   polyketides can be produced. The host cells are useful for constructing
CC   a library, where each individual colony of the library represents a
CC   colony with the ability to produce a particular PKS synthase and
CC   ultimately a particular polyketide. The polyketides produced by these
CC   colonies can be used collectively in a panel to represent a library or
CC   may be assessed individually for activity. Colonies in the library are
CC   also induced to produce the relevant synthases and thus to produce the
CC   relevant polyketides to obtain a library of candidate polyketides which
CC   can be screened for binding to desired targets such as receptors,
CC   signalling proteins, etc. The present sequence represents the DNA
CC   sequence of cosmid pKOS28-26 which encodes one or more domains of
CC   S. cellulosum PKS.
CC   Note: The present sequence is said to encode the functional domains
CC   of S. cellulosum PKS which correspond to domains or domain subsets of
CC   the amino acid sequences of ORF1 (AAU10700) and ORF2 (AAU10701).
XX
SQ   Sequence 33529 BP; 4489 A; 9518 C; 14470 G; 5046 T; 6 other;

  Query Match             43.0%;  Score 43.4;  DB 23;  Length 33529;
  Best Local Similarity   64.4%;  Pred. No. 0.095;
  Matches   65;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           |||  ||||  |||| |    || |  ||    |||||||  ||| | |||  | || |
Db   22758 CCGATGCAGACGGTGCTGAGCGGGGACGAAGCGGCGGTGCTCGCGGTGGCGCGGCGGCTG 22817

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
```

```
          | |||||  |||  ||||||   ||| |||| ||| |||||
Db  22818 GAGGCGCAGGGCCGGCGCACGCGGCGTCTGCGTGTGTCGCA 22858


RESULT 8
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match           43.0%;  Score 43.4;  DB 22;  Length 65140;
  Best Local Similarity   66.7%;  Pred. No. 0.092;
  Matches   62;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCG 68
           ||| ||  |||| || |  ||| | || |||| |    |  | ||  | ||||| |  ||
Db    4848 GGTCGTCGTCGCCGGAGCCGAGGACGCCGTGCGGCAACTGACCGCCCGCTTCGCCGACCG 4907

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             ||||  ||||||||  ||||||   ||||||||
Db    4908 CGGCCGGCGCACCAGCCGGCTGGCCGTCTCGCA 4940


RESULT 9
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
```

```
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             43.0%;  Score 43.4;  DB 22;  Length 125401;
  Best Local Similarity   66.7%;  Pred. No. 0.089;
  Matches   62;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGGTTCGCGGCGCG 68
           ||| ||  |||| || |  ||| | || |||| |    |  | ||  | ||||| |  ||
Db   65109 GGTCGTCGTCGCCGGAGCCGAGGACGCCGTGCGGCAACTGACCGCCCGCTTCGCCGACCG 65168

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           ||||  ||||||||  ||||||   ||||||||
Db   65169 CGGCCGGCGCACCAGCCGGCTGGCCGTCTCGCA 65201


RESULT 10
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             41.8%;  Score 42.2;  DB 22;  Length 27541;
```

```
  Best Local Similarity   65.3%;  Pred. No. 0.18;
  Matches   62;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||| ||   |||||||||| | || |  ||| | || |   |  ||||||| ||   ||||
Db  13838 CCGCAGTCCGTGGTGATCTCCGGTGACGAGGAGGCCGCCGAAACGATCGCCGCCACGTTC 13897

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGT 95
          || |  |||||       ||||||||||||  ||
Db  13898 GCCGAACGCGGGCGCAAGACCAAGCGGCTGCGGGT 13932


RESULT 11
AAH52062
ID   AAH52062 standard; DNA; 4851 BP.
XX
AC   AAH52062;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Mycobacterium tuberculosis potential drug target gene SEQ ID 116.
XX
KW   Drug target; growth; organism viability; characterisation; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200135317-A1.
XX
PD   17-MAY-2001.
XX
PF   13-NOV-2000; 2000WO-US31152.
XX
PR   12-NOV-1999;   99US-0165086.
PR   12-NOV-1999;   99US-0165124.
PR   01-FEB-2000; 2000US-0179531.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Eisenberg D,  Rotstein SH,  Marcotte EM;
XX
DR   WPI; 2001-329193/34.
DR   P-PSDB; AAG81211.
XX
PT   Identifying nucleotide or polypeptide sequence for use as drug target,
PT   involves providing algorithm that analyzes a functional relationship
PT   between nucleotide or polypeptide sequences, and comparing the
PT   sequences  -
XX
PS   Disclosure; Page 131-133; 207pp; English.
XX
CC   This invention relates to a method for identifying a nucleotide or
CC   polypeptide sequence that may be a drug target, or essential for growth
CC   or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC   represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC   tuberculosis proteins which are potential drug targets. The DNA and
CC   protein sequences are used to illustrate the method of the invention. The
CC   method involves providing an unknown nucleotide or polypeptide sequences,
CC   and comparing it to a number of sequences along with at least one
CC   algorithm capable of analysing a functional relationship between
CC   nucleotide and polypeptide sequences. The method is useful for
CC   characterising the function of nucleic acids and polypeptides that may be
CC   useful as a target for a drug or essential for the growth or viability of
CC   an organism.
XX
SQ   Sequence 4851 BP; 611 A; 1403 C; 1914 G; 923 T; 0 other;

  Query Match             41.4%;  Score 41.8;  DB 22;  Length 4851;
  Best Local Similarity   63.4%;  Pred. No. 0.24;
  Matches   64;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          || ||   |||||||||| | || | | ||   ||||   | ||||| ||||   |||||
Db    712 CCCGAATCGGTGGTGATCTCCGGTGCGCAGGCCGCGGCAAATGCGATTGCGGATCGGTTC 771

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          || ||||  ||   |||     | | | ||   ||||||||
Db    772 GCCGCGCAGGGTCGGCGGGTGCACCAGTTGGCGGTCTCGCA 812


RESULT 12
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             39.8%;  Score 40.2;  DB 22;  Length 31422;
  Best Local Similarity   64.5%;  Pred. No. 0.52;
  Matches   60;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
          |||||||  || | ||||  ||| | | ||||  |   || ||||   ||||||| | |||
Db  10284 GGTGGTGTTCTCCGGCGCCGAGGACGAGGTGGGGAACATGGCGGACTGGTTCGCCGAGCG 10343

Qy     69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||   | |   |||||| ||||    |  |||
```

```
Db  10344 CGGGCGGAGAGTCAAGCGCCTGCGAACCGGGCA 10376


RESULT 13
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             38.2%;  Score 38.6;  DB 21;  Length 31422;
  Best Local Similarity   63.4%;  Pred. No. 1.2;
  Matches   59;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||  | | ||||  ||| | | ||||  |   || ||||   ||||||| | |||
Db  10284 GGTGGTGTCCTCCGGCGCCGAGGACGAGGTGGGGAACATGGCGGACTGGTTCGCCGAGCG 10343

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   | |   |||||| ||||    |  |||
Db  10344 CGGGCGGAGAGTCAAGCGCCTGCGAACCGGGCA 10376


RESULT 14
AAQ10212/c
ID   AAQ10212 standard; DNA; 2721 BP.
XX
AC   AAQ10212;
XX
DT   17-DEC-2001  (updated)
DT   27-MAR-1991  (first entry)
XX
DE   BamHI J-I fragment carrying sequences characteristic of productive
DE   pseudorabies virus.
XX
KW   PRV; ss.
XX
OS   Pseudorabies virus.
XX
PN   USN7537855-N.
XX
PD   18-DEC-1990.
XX
PF   13-JUN-1990;   90US-0238940.
XX
PR   13-JUN-1990;   90US-0537855.
XX
PA   (USDA ) US AGRIC RES SERV.
XX
PI   Cheung AK;
XX
DR   WPI; 1991-021957/03.
XX
PT   Pseudo-rabies virus nucleotide sequences - used for producing
PT   nucleic acid probes, antigens and antibodies for distinguishing
PT   latent from productive infection
XX
PS   Disclosure; Page 21; 27pp; English.
XX
CC   The fragment carries sequences characteristic of the productive
CC   pseudorabies viral genome, and may be used as a probe in diagnosis
CC   of infection.
CC   (Note: Revised entry submitted to correct the patent number format of
CC   US Government-owned NTIS applications to prevent clashes with ongoing US
CC   granted patent numbers. For further information please visit the Derwent
CC   web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ   Sequence 2721 BP; 428 A; 1007 C; 1017 G; 269 T; 0 other;

  Query Match             37.8%;  Score 38.2;  DB 12;  Length 2721;
  Best Local Similarity   61.6%;  Pred. No. 1.7;
  Matches   61;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
             ||| |  || || | | ||||  |||| | ||||||  |  ||||  |||||||| |  |
Db     787 CGGCGGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGGGCT 728

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
           |||||  ||  |  ||  ||   |  || |   || |||
Db     727 CGGCGGCCGAGGATCGTCCCGGTCCCCTTCTCCTCCCGC 689


RESULT 15
AAQ10543/c
ID   AAQ10543 standard; DNA; 2721 BP.
XX
AC   AAQ10543;
XX
DT   17-DEC-2001  (updated)
DT   27-MAR-1991  (first entry)
XX
DE   BamHI J-I fragment carrying sequences characteristic of latent
DE   pseudorabies virus.
XX
```

KW PRV; ss.
XX
OS Pseudorabies virus.
XX
PN USN7537855-N.
XX
PD 18-DEC-1990.
XX
PF 13-JUN-1990; 90US-0238940.
XX
PR 13-JUN-1990; 90US-0537855.
XX
PA (USDA) US AGRIC RES SERV.
XX
PI Cheung AK;
XX
DR WPI; 1991-021957/03.
XX
PT Pseudo-raabies nucleotide sequences - used for producing
PT nucleic acid probes, antigens and antibodies for distinguishing
PT latent from productive infection
XX
PS Disclosure; Page 22; 27pp; English.
XX
CC The fragment carries sequences characteristic of the latent
CC pseudorabies viral genome, and may be used as a probe in diagnosis
CC of infection.
CC (Note: Revised entry submitted to correct the patent number format of
CC US Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
XX
SQ Sequence 2721 BP; 428 A; 1007 C; 1017 G; 269 T; 0 other;

Query Match 37.8%; Score 38.2; DB 12; Length 2721;
Best Local Similarity 61.6%; Pred. No. 1.7;
Matches 61; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

```
Qy       2 CGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCG 61
           ||| |  || || | | ||||  |||| | ||||||  |  ||||  ||||||||  |  |
Db     787 CGGCGGCGGCGGAGGTGGCGGCGGTGGTGGAAGCGGCGGCGGCGGCCGCGGCGGAGGGCT 728

Qy      62 CGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGC 100
           |||||  ||  |  ||  ||   |  || |   || |||
Db     727 CGGCGGCCGAGGATCGTCCCGGTCCCCTTCTCCTCCCGC 689
```

Search completed: November 5, 2002, 12:33:56
Job time : 242.215 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 458.785 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_18300_18400
Perfect score: 101
Sequence: 1 ccggagcaggtggtgatcgc.....aagcggctgcatgtctccca 101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		101	100.0	58733	1	AF217189	AF217189 Sorangium
2		101	100.0	71989	6	AR172664	AR172664 Sequence
3		99.4	98.4	68750	1	AF210843	AF210843 Sorangium
4		99.4	98.4	68750	6	AR193029	AR193029 Sequence
5		99.4	98.4	68750	6	AR199551	AR199551 Sequence
6		99.4	98.4	68750	6	AR199559	AR199559 Sequence
7		99.4	98.4	68750	6	AR199567	AR199567 Sequence
8		99.4	98.4	68750	6	AR201097	AR201097 Sequence
9		99.4	98.4	68750	6	AR208671	AR208671 Sequence
10		77	76.2	16124	1	AX024384	AX024384 Sequence
11		77	76.2	16124	6	AX024277	AX024277 Sequence
12		51.4	50.9	20394	1	SNA132222	AJ132222 Streptomy
13		51.4	50.9	20394	6	AX067996	AX067996 Sequence
14		51.4	50.9	43280	1	SFU78289	U78289 Streptomyce
15	c	51.4	50.9	84985	1	SNA278573	AJ278573 Streptomy
16		49.8	49.3	3900	1	AF262754	AF262754 Amycolato
17	c	48.2	47.7	26195	1	SC1G7	AL591083 Streptomy
18	c	45	44.6	39314	1	SGR300302	AJ300302 Streptomy
19		43.4	43.0	33529	6	AR166425	AR166425 Sequence
20		43.4	43.0	65140	6	AX211705	AX211705 Sequence
21		43.4	43.0	123580	1	AF263912	AF263912 Streptomy
22		43.4	43.0	125401	6	AX211739	AX211739 Sequence
23		42.4	42.0	113193	1	AF357202	AF357202 Streptomy
24		42.2	41.8	27541	6	AX211706	AX211706 Sequence
25	c	41.8	41.4	16096	1	AE007124	AE007124 Mycobacte
26	c	41.8	41.4	24292	1	MTCY24G1	Z83858 Mycobacteri
27		41.8	41.4	31175	1	MTU00024	U00024 Mycobacteri
28		41.8	41.4	31176	1	MSGY2	AD000009 Mycobacte
29	c	40.2	39.8	18023	1	AX024392	AX024392 Sequence
30	c	40.2	39.8	18023	6	AX024285	AX024285 Sequence
31		40.2	39.8	25150	1	AB070945	AB070945 Streptomy
32		40.2	39.8	31422	6	E38021	E38021 Avermectin
33	c	40.2	39.8	64957	1	AB032367	AB032367 Streptomy
34		40.2	39.8	78210	1	AB070949	AB070949 Streptomy
35		40.2	39.8	344050	1	MLEPRTN1	AL583917 Mycobacte
36	c	39.6	39.2	24225	1	SC2C4	AL512902 Streptomy
37		38.4	38.0	168764	8	AC087726	AC087726 Chlamydom
38		38.2	37.8	217	14	SH1EGB	M25527 Pseudorabie
39	c	38.2	37.8	2080	14	SH1ULTRJ	L24958 Pseudorabie
40	c	38.2	37.8	4968	14	SH1SEQ	M81222 Herpesvirus
41		38.2	37.8	5123	14	HEPVIE	X15120 Pseudorabie
42	c	38.2	37.8	8438	14	SH1LLT	M57505 Pseudorabie
43	c	38.2	37.8	175911	2	AC129453	AC129453 Rattus no
44		37.8	37.4	104326	1	AB070940	AB070940 Streptomy
45		37.6	37.2	77534	1	AF235504	AF235504 Streptomy

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

AUTHORS Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and Julien,B.
TITLE Cloning and heterologous expression of the epothilone gene cluster
JOURNAL Science 287 (5453), 640-642 (2000)
MEDLINE 20115953
PUBMED 10649995
REFERENCE 2 (bases 1 to 58733)
AUTHORS Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
TITLE Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum
JOURNAL Gene 249 (1-2), 153-160 (2000)
MEDLINE 20293058
PUBMED 10831849
REFERENCE 3 (bases 1 to 58733)
AUTHORS Julien,B.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1. .58733
/organism="Polyangium cellulosum"
/strain="SMP44"
/db_xref="taxon:56"

CDS <1. .992
/note="orfA"
/codon_start=3
/transl_table=11
/product="putative transposase"
/protein_id="AAF62878.1"
/db_xref="GI:7453555"
/translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRAREHR
GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
HIGAVRQVIDRRRRSERHLFPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"

CDS 989. .1501
/note="orfB"
/codon_start=1
/transl_table=11
/product="putative transposase"
/protein_id="AAF62879.1"
/db_xref="GI:7453556"
/translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
RTKQRRTKKH"

gene 1998. .6263
/gene="epoA"

CDS 1998. .6263
/gene="epoA"
/note="type I polyketide synthase; epothilone PKS loading module"
/codon_start=1
/transl_table=11
/product="EpoA"
/protein_id="AAF62880.1"
/db_xref="GI:7453557"
/translation="MADRPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"

gene 6260. .10492
/gene="epoB"

CDS 6260. .10492
/gene="epoB"
/note="non ribosomal peptide synthetase; epothilone PKS module 1"
/codon_start=1
/transl_table=11
/product="EpoB"
/protein_id="AAF62881.1"
/db_xref="GI:7453558"
/translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"

gene 10639. .16137
/gene="epoC"

CDS 10639. .16137
/gene="epoC"
/note="type I polyketide synthase; epothilone PKS module 2"
/codon_start=1
/transl_table=11
/product="EpoC"
/protein_id="AAF62882.1"
/db_xref="GI:7453559"
/translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match             100.0%;  Score 101;  DB 1;  Length 58733;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18300 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 18359

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   18360 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 18400


RESULT 2
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others

ORIGIN

  Query Match             100.0%;  Score 101;  DB 6;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-10;
  Matches  101;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   18300 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 18359

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||||||||||||||||||||||||||||||||||||||||
Db   18360 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 18400


RESULT 3
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
```

```
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
CDS                  3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
CDS                  complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVTVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
CDS                  6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVLTGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRIAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYAFFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene                 7610. .11875
                     /gene="epoA"
CDS                  7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRFIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature         7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
misc_feature         9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
misc_feature         10529. .11428
                     /gene="epoA"
                     /note="Region: enoyl reductase"
misc_feature         11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
gene                 11872. .16104
                     /gene="epoP"
CDS                  11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene                 16251. .21749
                     /gene="epoB"
CDS                  16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

Query Match          98.4%;  Score 99.4;  DB 1;  Length 68750;
```

```
  Best Local Similarity   99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 4
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 5
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 6
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 7
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity   99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012
```

```
RESULT 8
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity  99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 9
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match            98.4%;  Score 99.4;  DB 6;  Length 68750;
  Best Local Similarity  99.0%;  Pred. No. 9.3e-10;
  Matches  100;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23912 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 23971

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||||||||||| |||||||||||||||||||||||||||
Db  23972 GCGGCGCGCGGCGCGCGCACCAAGCGGCTGCATGTCTCGCA 24012


RESULT 10
AX024384
LOCUS       AX024384               16124 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024384
VERSION     AX024384.1  GI:10184588
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match            76.2%;  Score 77;  DB 1;  Length 16124;
  Best Local Similarity  85.1%;  Pred. No. 1.7e-05;
  Matches   86;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          |||||||||||||||||||||||||  ||| ||   ||||||  |||||||||||| ||||
Db   1797 CCGGAGCAGGTGGTGATCGCGGGCGCCGAGAAATTCGTGCAGCAGATCGCGGCGGCGTTC 1856

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||| || | ||| ||||| | ||||||||||||||||
Db   1857 GCGGCGCGGGGGGCGCGAACCAAACCGCTGCATGTCTCGCA 1897


RESULT 11
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match            76.2%;  Score 77;  DB 6;  Length 16124;
  Best Local Similarity  85.1%;  Pred. No. 1.7e-05;
  Matches   86;  Conservative    0;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          |||||||||||||||||||||||||  ||| ||   ||||||  |||||||||||| ||||
Db   1797 CCGGAGCAGGTGGTGATCGCGGGCGCCGAGAAATTCGTGCAGCAGATCGCGGCGGCGTTC 1856

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          |||||||| || | ||| ||||| | ||||||||||||||||
Db   1857 GCGGCGCGGGGGGCGCGAACCAAACCGCTGCATGTCTCGCA 1897


RESULT 12
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
```

```
KEYWORDS    pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
  TITLE     The biosynthetic gene cluster for the 26-membered ring polyene
            macrolide pimaricin. A new polyketide synthase organization encoded
            by two subclusters separated by functionalization genes
  JOURNAL   J. Biol. Chem. 274 (15), 10133-10139 (1999)
  MEDLINE   99214571
   PUBMED   10187796
REFERENCE   2  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.
            Real, 1-Leon, 24006, SPAIN
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     gene            1. .20394
                     /gene="pimS1"
     CDS             1. .20394
                     /gene="pimS1"
                     /function="pimaricin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="CAB41041.1"
                     /db_xref="GI:4678703"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
                     HAFHSPLMEPMLADFRAVSETITFGTPTAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
```

```
                    LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                    DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                    VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                    DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT      2611 a    7486 c    7397 g    2900 t
ORIGIN

  Query Match             50.9%;  Score 51.4;  DB 1;  Length 20394;
  Best Local Similarity   69.3%;  Pred. No. 0.88;
  Matches    70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           |||  |  ||| |||||| ||||||  ||| | ||||||| ||| |||||||||   |||
Db   11944 CCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACGCGGTGCTGGCCATCGCGGCGCACTTC 12003

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||| |   ||       ||||  || ||||  ||||||||
Db   12004 GCGGGGGAGGGGCGCAAAACCACCCGACTGCGGGTCTCGCA 12044


RESULT 13
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24642.1"
                     /db_xref="GI:12329807"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVIKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAFEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAFPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVFFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTECVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEACFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGREPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDREEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVFPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVIDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRFAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLFDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
```

```
                     HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                     LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                     DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                     VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                     DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a   7486 c   7397 g   2900 t
ORIGIN

  Query Match           50.9%;  Score 51.4;  DB 6;  Length 20394;
  Best Local Similarity   69.3%;  Pred. No. 0.88;
  Matches   70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      1 CCGGAGCAGGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
          |||  |  ||| ||||||  ||||||  ||| | ||||||| ||| |||||||||   |||
Db  11944 CCGTCGTCGGTCGTGATCTCGGGCGCCGAGGACGCGGTGCTGGCCATCGCGGCGCACTTC 12003

Qy     61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
          ||||  |   ||       ||||  || ||||  ||||||||
Db  12004 GCGGGGGAGGGGCGCAAAACCACCCGACTGCGGGTCTCGCA 12044


RESULT 14
SFU78289
LOCUS       SFU78289               43280 bp    DNA     linear   BCT 13-AUG-1997
DEFINITION  Streptomyces fradiae tylactone synthase, starter module and modules
            1-7, (tylG) gene, complete cds.
ACCESSION   U78289
VERSION     U78289.1  GI:2317859
KEYWORDS    .
SOURCE      Streptomyces fradiae.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Sequence of Streptomyces fradiae tylactone synthase gene tylG
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
            Center, Indianapolis, IN 46285, USA
COMMENT     tylG.
FEATURES             Location/Qualifiers
     source          1. .43280
                     /organism="Streptomyces fradiae"
                     /db_xref="taxon:1906"
     gene            1. .43280
                     /gene="tylG"
     CDS             816. .14234
                     /gene="tylG"
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tylactone synthase starter module and modules 1
                     & 2"
                     /protein_id="AAB66504.1"
                     /db_xref="GI:2317860"
                     /translation="MSSALRRAVQSNCGYGDLMTSNTAAQNTGDQEDVDGPDSTHGGE
                     IAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWRAALEDHAGFDAGFFGMNAR
                     QAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYAVLTRRSAVSAG
                     GYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGETSLAVAG
                     GVNLILTEESTTVMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAALADGD
                     RVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTGTRA
                     GDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGEL
                     VPSLNFSTPNPDIPLDDLRLRVQTERQEWNEEDDRPRVAGVSSFGMGGTNVHLVIAEA
                     PAAAGSSGAGGSGAGSGAGISAVSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLA
                     DVAVTMADRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVLDGGVVVGAAP
                     GGAGAAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSV
                     HVGWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAA
                     TVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVN
                     GPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRP
                     RSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHP
                     VLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL
                     TTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR
                     HTAVVLGLDTAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTPRA
                     LAQFLHTELLDGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSPEALWRLVVEG
                     VDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGISPREAVAM
                     DPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRLAEGAEGSDGYLLT
                     GTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMALAGGVTVM
                     AGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAV
                     VCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGDPIE
                     AQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVD
                     EPSRHVDWSAGGVWLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAPDTAEAESA
                     TTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVT
                     MAGRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVF
                     PGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVSGGAGLERVDVVQP
                     VTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGR
                     YLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVNGPASTVVSGDRRAVAGYVAVCQAE
                     GVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAG
                     YWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLR
                     RQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHLTTLPTYPFNHHHYWLDTTPTTPAT
                     TTQSPTDAWRYRVTWKALTESSPVRPHSIGRCLLVAPPTTDGELLDGLTTVLSERGAS
                     VARLEVPIGARRAEVAELLKPSMESAGEENTTVVSLLGLVPSTDAVRTSIALLQAVSD
                     IGVPAARVWALTRRAVAVVPGETPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAEL
                     TRTPETSQPPQTPERLPQTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAA
                     AGAASWQPSGTVLITGGMGAIGRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELR
                     GHGCEVVHAACDVAERDALAAALVTAYPPNAVFHTAGILDDAVIDTLSPESFETVRGAK
                     VCGAELLHQLTADIKGLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLP
                     ATSVAWGLWGGGGMAAGAGEESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVDW
                     ERFAPATNAIRPGRLFDTVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKL
                     VELVRTEAAFVLRHPNTDAIGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFD
                     HPSPTALARLLLDRLTGAGAPAPAADEPPLPVAVADDDPVVIVGMACRFPGGAGTPEA
                     LWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGI
                     SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPE
                     EYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMA
                     LAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARR
                     LGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVDVVEAHGTG
                     TRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGW
                     LPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAP
                     EALEALEATDAPEAPEAPEAPDVTDVTEALEAPDATEAEGAKAPGSPEEAQPAVGVVP
                     VVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVTMAGRSRFGYRAVVLARGEAELAG
                     RLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAA
                     SMRECARALSVHVEWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAV
                     VGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA
                     KWPGVQVAAVNGPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKG
                     ELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEQG
                     HRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATL
                     TWDPALPPGHLTTLPTYPFNHHHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLA
                     DSGERVFAGRLAGSEHDWLTDHAVSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEA
                     PLVLPAAGGVRVQMRVSAADESGRRRVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL
                     VPDPEPTPPDADWARAWPPAGERVEPAELYERFGALGYEYGEAFAGVRAVWRQPDALL
                     AEVLLPDRASTGAGRFGVHPALLDAALQPWIAGGLLEVPEDAVLLPFAWQGVSLYATG
                     AGALRVRLTKAGDGAVSLQAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGERSLY
                     RVEWQPRLLPAGPPRSWAVLGPDADRLAGTPGLGDQPDGGPTALYPEVRALRKALAAG
                     APRPEAVVLPVLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAV
                     PGEEIGDLACAGVWGLVRSARSEHPGRFALVDTDGHPDDRTALPLALRAVLDGAGQLS
                     LRAGTARTPVLLRAGTPEEQRGPAFDPAGTVLVTGATGTLGRLLARHLAAEHGVRHLL
                     LLSRGGRAAEGADELAAELAGLEAEPCFAACDAADREALARVLAEVPADRPLTGVIHA
                     AGVLDDGTLDALTPERIGTVMRPKADAALNLHELTRTSPLSVFAVFSGAAGILGRPGQ
                     ANYAAANTFLDALAQHRRAHGLPAVSLAWGLWGGATGMTGHLSGTDLRRMRRSGIAPM
                     THDQGLALFDRALAASAEDPLLVPMRLDLAALVRERAEHGPDAVPGPLLGLLPARAAV
```

```
                RQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEP
                GRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRDGG
                DTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAALTPQGRNPSAPAPAPSDL
                DERLDSANDDDLFAFIEEQL"
misc_feature    942. .3929
                /gene="tylG"
                /function="encodes starter module"
misc_feature    942. .2156
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase KS-Q of
                starter module"
misc_feature    2571. .3557
                /gene="tylG"
                /function="encodes acyltransferase domain of starter
                module"
misc_feature    3675. .3929
                /gene="tylG"
                /function="encodes acyl carrier protein domain of starter
                module"
misc_feature    3993. .8471
                /gene="tylG"
                /function="encodes module 1"
misc_feature    3993. .5264
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase of module 1"
misc_feature    5631. .6617
                /gene="tylG"
                /function="encodes acyltransferase domain of module 1"
misc_feature    7410. .7949
                /gene="tylG"
                /function="encodes beta-ketoreductase domain of module 1"
misc_feature    8220. .8471
                /gene="tylG"
                /function="encodes acyl carrier protein domain of module
                1"
misc_feature    8541. .13970
                /gene="tylG"
                /function="encodes module 2"
misc_feature    8541. .9812
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase of module 2"
misc_feature    10260. .11246
                /gene="tylG"
                /function="encodes acyltransferase domain of module 2"
misc_feature    11319. .11876
                /gene="tylG"
                /function="encodes dehydratase domain of module 2"
misc_feature    12861. .13415
                /gene="tylG"
                /function="encodes beta-ketoreductase domain of module 2"
misc_feature    13719. .13970
                /gene="tylG"
                /function="encodes acyl carrier protein domain of module
                2"
CDS             14351. .19945
                /gene="tylG"
                /note="ORF2"
                /codon_start=1
                /transl_table=11
                /product="tylactone synthase module 3"
                /protein_id="AAB66505.1"
                /db_xref="GI:2317861"
                /translation="MTAELVATRKRLGALEERAREPIAVVAMSCRYPGGVTTPEDLWR
                LLADERDAVSGLPRDRGWDLDALYDPDGGPGTSYAREGGFLSHCAGFDAEFFGISPRE
                ALAMDPQQRLLLETSWEALERAGVTADRARGSRTGVYAGVMYDDYGARVLYGAGAGPP
                EDLEGYLVNGSAGSIASGRVSYTFGLRGPAVTVNTACSSSLVSLHLAVRALRNGECDM
                ALAGGATVLSTPTVLVDFSRQRGLAPDGRCKAFADSADGTSWAEGAGMLLLQRLSDAR
                REGRPVLAVIRGSAVNQDGASNGLTAPNGRAQRQVIEDALRDAGVGPDQVDAVEAHGT
                GTELGDPIEAGALLATYGQARTAERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMRHG
                RLPRTLHVDRPTTRVDWEKGGVRLLTEPVPWPGEAGEPRRAGVSSFGASGTNAHVVLE
                SVPAGEPPAAGRPEDTGGAWTVSGRGPAALRAQAARLYDALTGTGTGTGQGAGQGAGP
                GTAEVAGALAHARTAFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERR
                TAFLFSGQGSQRAGSGRGLYRRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEA
                EPLDRTEFTQPALFALQTALFRLAEHHGLRAEALCGHSVGEIAAAHAAGVLTLPDAAR
                LVAARGRLMQALPAGGAMAALRATAEEIAPLLERRAGELALAAVNGPSSVVVSGDEAA
                VLELLEQWRAEGREARRLAVSHAFHSPRMDGMLTQFDRVARTLTFAPPTIPLVSTLTG
                TPVTEETLCTADHWVRQAREPVRFLDAMRTLRADGIDTFVELGPDGVLSAMARDCADD
                RPDGDTTGAGDGETPDPLLTLPLLRRSVPETGDAEHPGGFERALATAYAHGVPLRLAP
                APDAASLAVAAELPTYAFQRTHYWLDAPAAPAALPAGLDDAGHPLLSAALDLPGGRGT
                VWTGALSAATLPWAADHSVHGRTVLPGTALLDLALHAAPRVGELTFEAPLVLPEDGEV
                RLRVVLAEPDASGVRELSVHSAGEDGGWTRHATAVLDTGTTTAGEPAGAPPAAWPPGD
                AEPLDLAAEYERFADAGIGYGPAFRGLRSAWRDGDAILADVRLPGELAGEADRYGIHP
                ALLDAALHTAAAALGGAHGMLPFTWNGVTLHARGAHAIRVRLTPAGPDAVAVTAVDPA
                GRPVFTAASLTLRPVTTGQLTAAEAARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDL
                LPGETPHHPDLASALADTGTAPFRVLADLRGYGTATPRELASQALALVQQWADAAEAA
                EGRLVLVTRRAVDIGDGVTDPAAATVWGLVRAAQSEHPGCFALLDTDDSPRSRQLLPR
                VAGTAEQLALRDGTLLAPSLTRATLPAGARLPALDGTVLITGGTGSLGAEAARHLVTR
                HGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAACDVADRAALRALLDRVPAGHP
                LTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLHELTQGHALSAFILYSSAAG
                VLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPWDGGGMASGLGGTDTARLRR
                SGIAPLSRAEGLAALDAALAAGGDDTAPAHLLPIRVDAVTLRGADTVPAVLRDLAGTA
```

Query Match 50.9%; Score 51.4; DB 1; Length 43280;
Best Local Similarity 72.0%; Pred. No. 0.76;
Matches 67; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

```
Qy       9 GGTGGTGATCGCGGGCGTGGAGCAAGCGGTGCAGGCGATCGCGGCGGGGTTCGCGGCGCG 68
           |||||||||| | |||| |||| ||||||||   | |   ||||||  | |||| |  ||
Db   38468 GGTGGTGATCTCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACG 38527

Qy      69 CGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
             |||   ||||||| |||||| |  ||| ||||
Db   38528 AGGCCGCCGCACCAGGCGGCTCCGCGTCGCGCA 38560
```

```
RESULT 15
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
```

```
                /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                ALADRDPGRGQV"
gene            2190. .3566
                /gene="pimK"
CDS             2190. .3566
                /gene="pimK"
                /function="putative glucosyl transferase"
                /codon_start=1
                /transl_table=11
                /product="PimK protein"
                /protein_id="CAC20918.1"
                /db_xref="GI:12055069"
                /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAAIMRSAGGRETAAGL
                VLSLPALA"
gene            complement(3662. .9736)
                /gene="pimS4"
CDS             complement(3662. .9736)
                /gene="pimS4"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS4 protein"
                /protein_id="CAC20919.1"
                /db_xref="GI:12055070"
                /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772. .15198)
                /gene="pimS3"
CDS             complement(9772. .15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
```

```
                    PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                    VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                    SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                    GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                    LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                    RTLPDEQIDAAHSLVRDAIFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                    VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                    VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                    LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                    AERGVRHLLLVSRRGADAPGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                    HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                    SGTVGAAGQGNYAAANAFIDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                    LARMGTPALTEEQGIALFIAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                    RSVVSASETAVTLVQQLSFLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                    GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match             50.9%;  Score 51.4;  DB 1;  Length 84985;
  Best Local Similarity   69.3%;  Pred. No. 0.66;
  Matches   70;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy       1 CCGGAGCAGGTGGTGATCGCGGGCGTGGACCAAGCGGTGCAGGCGATCGCGGCGGGGTTC 60
           |||  |  ||| ||||||| |||||||  ||| | ||||||| ||| |||||||||   |||
Db   69209 CCGTCGTCGGTCGTGATCTCGGGCGCCGACGACGCGGTGCTGGCCATCGCGGCGCACTTC 69150

Qy      61 GCGGCGCGCGGCGTGCGCACCAAGCGGCTGCATGTCTCGCA 101
           |||| |    ||       ||||  || ||||  ||||||||
Db   69149 GCGGGGGAGGGGCGCAAAACCACCCGACTGCGGGTCTCGCA 69109


Search completed: November  5, 2002, 12:08:15
Job time : 630.785 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 1003.2 Seconds
(without alignments)
1469.095 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065
Perfect score: 91
Sequence: 1 ccgacgcgcaacgggatggc.....tcgaccgggttgaccgcgcc 91

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	34	37.4	577	17	AQ162259	AQ162259 mgxb0012K
c	2	30.6	33.6	1160	12	BG844853	BG844853 1024008A0
	3	30.2	33.2	817	12	BG343906	BG343906 HVSMEg000
c	4	29.8	32.7	553	12	BE725325	BE725325 894082D06
c	5	29.8	32.7	603	13	BI530795	BI530795 1024108G1
c	6	29.8	32.7	668	13	BI530912	BI530912 1024109F0
	7	29.6	32.5	560	17	AZ935138	AZ935138 BJ__Ba000
	8	29.6	32.5	656	17	AZ935404	AZ935404 BJ__Ba000
	9	29.6	32.5	658	14	BQ765891	BQ765891 EBro03_SQ
c	10	29.4	32.3	488	14	BM691860	BM691860 UI-E-CL1-
	11	29.4	32.3	623	17	BH807613	BH807613 1008072D0
c	12	29	31.9	398	10	BB739356	BB739356 BB739356
	13	29	31.9	523	10	AV618984	AV618984 AV618984
c	14	29	31.9	549	10	BE219908	BE219908 hv64d03.x
	15	28.6	31.4	224	9	AI920329	AI920329 603019C06
c	16	28.6	31.4	510	12	BG897901	BG897901 HOA1-1-H6
	17	28.6	31.4	561	9	AU148928	AU148928 AU148928
c	18	28.4	31.2	387	12	BG343971	BG343971 HVSMEg000
c	19	28.4	31.2	401	17	AQ909025	AQ909025 GSSTc0987
c	20	28.2	31.0	219	14	BQ322897	BQ322897 RC0-CS004
	21	28.2	31.0	605	10	AW286258	AW286258 LG1_329_D
	22	28.2	31.0	679	10	BB594172	BB594172 BB594172
c	23	28.2	31.0	874	12	BF254077	BF254077 HVSMEf000
c	24	28.2	31.0	881	17	CNS03K1K	AL247601 Tetraodon
c	25	28.2	31.0	1026	17	CNS04LVH	AL296630 Tetraodon
c	26	28	30.8	421	13	BJ264863	BJ264863 BJ264863
c	27	28	30.8	594	10	BE504110	BE504110 hv85d03.x
	28	27.8	30.5	469	10	BE583753	BE583753 7-4F-HA P
	29	27.8	30.5	512	9	AI588339	AI588339 fb99e09.x
	30	27.8	30.5	523	13	BI723385	BI723385 1031066H0
	31	27.8	30.5	530	10	BE583754	BE583754 2-11G-HA
	32	27.8	30.5	557	13	BI723387	BI723387 1031066H0
	33	27.8	30.5	601	13	BI723386	BI723386 1031066H0
	34	27.8	30.5	609	10	BE583577	BE583577 8-3H-HA P
	35	27.8	30.5	646	10	BE583752	BE583752 9-4E-HA P
c	36	27.8	30.5	965	9	AL532447	AL532447 AL532447
c	37	27.8	30.5	1057	17	CNS03BPC	AL236793 Tetraodon
	38	27.8	30.5	1064	14	BQ948399	BQ948399 AGENCOURT
c	39	27.8	30.5	1174	14	BQ876942	BQ876942 AGENCOURT
	40	27.6	30.3	224	17	BH221657	BH221657 1006102E0
	41	27.6	30.3	585	14	BQ034876	BQ034876 1091010B0
	42	27.6	30.3	666	10	BB653198	BB653198 BB653198
	43	27.6	30.3	1274	14	BQ065081	BQ065081 AGENCOURT
	44	27.4	30.1	268	14	BQ759911	BQ759911 EBpi05_SQ
	45	27.4	30.1	299	12	BF462534	BF462534 UI-M-CG0p

ALIGNMENTS

RESULT 1
AQ162259
LOCUS AQ162259 577 bp DNA linear GSS 09-SEP-1998
DEFINITION mgxb0012K23r CUGI Rice Blast BAC Library Magnaporthe grisea genomic clone mgxb0012K23r, DNA sequence.
ACCESSION AQ162259
VERSION AQ162259.1 GI:3558660
KEYWORDS GSS.
SOURCE Magnaporthe grisea.
ORGANISM Magnaporthe grisea
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
REFERENCE 1 (bases 1 to 577)
AUTHORS Yu,Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R., Phillips,K., Sasinowski,M, Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Dean RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson Universiy, Clemson, SC 29634
Tel: 864 656 5737
Fax: 864 656 4293
Email: rdean@clemson.edu
Seq primer: GGAAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 337.
FEATURES Location/Qualifiers

```
     source          1. .577
                     /organism="Magnaporthe grisea"
                     /strain="70-15"
                     /db_xref="taxon:148305"
                     /clone="mgxb0012F23r"
                     /clone_lib="CUGI Rice Blast BAC Library"
                     /tissue_type="Protoplasts"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pEACWICH; Site_1: HindIII; Site_2: HindIII;
                     Rice blast is one of the most devestating fungal diseases
                     of rice world wide. It is a filamentous ascomycete with
                     a haploid genome (n=7) of approximately 40 Mbp. Rice
                     blast is an important model fungal pathogen for studying
                     numerous aspects of the fungal-host interaction. In
                     order to facilitate genome wide analysis, a BAC library
                     containing 9216 clones with an average insert size of 130
                     kbp was constructed. This library represents greater
                     than 25X genome coverage. High density colony filters
                     are available upon request."
BASE COUNT      121 a    148 c    148 g    160 t
ORIGIN

  Query Match             37.4%;  Score 34;  DB 17;  Length 577;
  Best Local Similarity   61.1%;  Pred. No. 11;
  Matches   55;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           | | ||    ||||||   ||   |||| |||| | ||| |||  | |  ||||||||
Db     349 CCAGGCTGTGCGGGATAATGATACCATCCGGGCTGTTATCATGGGTACTGGTATCAACCA 408

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           || |||  | || | ||  | |||  |||
Db     409 AGACGGCAAGACGCCTGGTATCACCATGCC 438


RESULT 2
BG844853/c
LOCUS       BG844853                1160 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024008A08.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG844853
VERSION     BG844853.1  GI:14226037
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 1160)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .1160
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      81 a    493 c    525 g     35 t     26 others
ORIGIN

  Query Match             33.6%;  Score 30.6;  DB 12;  Length 1160;
  Best Local Similarity   59.3%;  Pred. No. 1.2e+02;
  Matches   51;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       5 CGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGA 64
           |||||  || |   |||  |||  | | ||||    |||||  | ||||||  || |||
Db     510 CGCGCGCCGCGCGCGCGCGCGCCGCGGCGCGCCCGCCCGGGCCCCGGCCAGCACCCACAC 451

Qy      65 TGGCCGGTCGACCGGGTTGACCGCGC 90
             ||   ||   |||  | ||||||
Db     450 GTNCCCCGCGCGGGGGCGGGCCGCGC 425


RESULT 3
BG343906
LOCUS       BG343906                817 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEg0007B09f Hordeum vulgare pre-anthesis spike EST library
            HVcDNA0008 (white to yellow anther) Hordeum vulgare cDNA clone
            HVSMEg0007B09f, mRNA sequence.
ACCESSION   BG343906
VERSION     BG343906.1  GI:13156235
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 817)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex pre-anthesis spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 418
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 6
            High quality sequence stop: 798.
FEATURES             Location/Qualifiers
     source          1. .817
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEg0007B09f"
                     /clone_lib="Hordeum vulgare pre-anthesis spike EST library
                     HVcDNA0008 (white to yellow anther)"
                     /tissue_type="pre-anthesis spike"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spike with awns trimmed were collected at white, green and
                     yellow anther stages (Fenton). Total RNA was prepared from
                     each pool, equal quantities of all three RNA pools were
                     combined, poly(A) RNA was purified from the mixture, one
```

primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close lab (Choi) at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing) Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch , Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 148 a 286 c 263 g 120 t
ORIGIN

Query Match 33.2%; Score 30.2; DB 12; Length 817;
Best Local Similarity 69.5%; Pred. No. 1.4e+02;
Matches 41; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

```
Qy      11 ACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGATGGCC 69
           |||||  |  |||| |||| ||||||| | ||||||   |||   |   | ||||||||
Db     653 ACGGGCCGAGGACCCCATCGGGGCGCTTAACCGGGGAGGGGCGTCCGGGCTTGATGGCC 711
```

RESULT 4
BE725325/c
LOCUS BE725325 553 bp mRNA linear EST 14-SEP-2000
DEFINITION 894082D06.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BE725325
VERSION BE725325.1 GI:10126605
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 553)
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2
JOURNAL Unpublished (2000)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .553
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 99 a 202 c 148 g 104 t
ORIGIN

Query Match 32.7%; Score 29.8; DB 12; Length 553;
Best Local Similarity 58.4%; Pred. No. 1.6e+02;
Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           | | |||||  |  |  |||  |||   | | ||||| |||| ||||  ||  | ||||
Db     418 CAAGGCGCACGGCAAGAGCGCTCGCGAGTCGACGCTGCTCCGCGGCTACGCGCTGAACCT 359

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGC 90
           ||  ||  | ||  ||  || ||  ||
Db     358 TGCCCGCGCGGCGCAGGGCATGCCCAAGC 330
```

RESULT 5
BI530795/c
LOCUS BI530795 603 bp mRNA linear EST 29-AUG-2001
DEFINITION 1024108G11.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI530795
VERSION BI530795.1 GI:15371369
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.
REFERENCE 1 (bases 1 to 603)
AUTHORS Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1024b
JOURNAL Unpublished (2001)
COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.
FEATURES Location/Qualifiers
source 1. .603
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (acetate-containing) medium in the light, TAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda ZAP clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 104 a 220 c 167 g 112 t
ORIGIN

Query Match 32.7%; Score 29.8; DB 13; Length 603;
Best Local Similarity 58.4%; Pred. No. 1.6e+02;
Matches 52; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```
Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
```

```
          | | |||||  |  |  |||  |||   , | ||||| |||| ||||  ||  | ||||
Db    418 CAAGGCGCACGGCAAGAGCGCTCGCGAGTCGACGCTGCTCCGCGGCTACGCGCTGAACCT 359

Qy     62 TGATGGCCGGTCGACCGGGTTGACCGCGC 90
          ||  ||  | ||   ||  || ||  ||
Db    358 TGCCCGCGCGGCGCAGGGCATGCCCAAGC 330


RESULT 6
BI530912/c
LOCUS       BI530912                668 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  1024109F09.x2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BI530912
VERSION     BI530912.1  GI:15371486
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
            ,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants. Project: 1024b
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Chlamydomonas reinhardtii"
                     /strain="CC-1690 wild type mt+ 21gr"
                     /db_xref="taxon:3055"
                     /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                     II"
                     /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                     XhoI; This library, constructed by John Davies and Jeffrey
                     McDermott, combines cDNAs from CC-1690 cells grown to
                     mid-log phase in TAP (acetate-containing) medium in the
                     light, TAP medium in the dark, HS (minimal) medium in
                     ambient levels of CO2 and HS medium bubbled with 5% CO2.
                     PolyA mRNA was purified from each sample, pooled and cDNA
                     synthesized. The cDNA was directionally cloned into lambda
                     ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                     pBluescript II SK- plasmids were excised from the lambda
                     ZAP clones by superinfection with ExAssist (Stratagene)
                     phage. The library was normalized using method 4 described
                     in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      116 a    236 c    184 g    132 t
ORIGIN

  Query Match           32.7%;  Score 29.8;  DB 13;  Length 668;
  Best Local Similarity  58.4%;  Pred. No. 1.7e+02;
  Matches   52;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy      2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
          | | |||||  |  |  |||  |||   | | ||||| |||| ||||  ||  | ||||
Db    418 CAAGGCGCACGGCAAGAGCGCTCGCGAGTCGACGCTGCTCCGCGGCTACGCGCTGAACCT 359

Qy     62 TGATGGCCGGTCGACCGGGTTGACCGCGC 90
          ||  ||  | ||   ||  || ||  ||
Db    358 TGCCCGCGCGGCGCAGGGCATGCCCAAGC 330


RESULT 7
AZ935138
LOCUS       AZ935138                560 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0003K18r B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ935138
VERSION     AZ935138.1  GI:13777370
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 525.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      117 a    164 c    172 g    107 t
ORIGIN

  Query Match           32.5%;  Score 29.6;  DB 17;  Length 560;
  Best Local Similarity  59.5%;  Pred. No. 1.8e+02;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      5 CGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGA 64
          ||   | | | |||||| || || |  |  |||| ||| ||| | ||  | ||| ||||
Db    342 CGACAATCAGCATGGCGCCCACAACAAGAAGCTGTTCCAGGGGTGGGTGAACAAGCATGT 401

Qy     65 TGGCCGGTCGACCGGGTTGACCGC 88
          ||  |   | | |  |  | ||||
Db    402 TGCGCTCGCCAACAAGGCGGCCGC 425


RESULT 8
AZ935404
LOCUS       AZ935404                656 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0003D08f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ935404
VERSION     AZ935404.1  GI:13778084
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
```

```
           Fax: 864 656 4293
           Email: rwing@clemson.edu
           Class: BAC ends
           High quality sequence stop: 605.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      138 a    193 c    195 g    130 t
ORIGIN

  Query Match             32.5%;  Score 29.6;  DB 17;  Length 656;
  Best Local Similarity   59.5%;  Pred. No. 1.9e+02;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      5 CGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGA 64
          ||   | | | |||||| || || |  |  |||| ||| ||| | ||  | ||| ||||
Db    342 CGACAATCAGCATGGCGCCCACAACAAGAAGCTGTTCCAGGGGTGGGTGAACAAGCATGT 401

Qy     65 TGGCCGGTCGACCGGGTTGACCGC 88
          ||  |   | | |  |  | ||||
Db    402 TGCGCTCGCCAACAAGGCGGCCGC 425


RESULT 9
BQ765891
LOCUS       BQ765891                 658 bp    mRNA    linear   EST 26-JUL-2002
DEFINITION  EBro03_SQ008_O11_R root, 3 week, waterlogged, cv Optic, EBro03
            Hordeum vulgare cDNA clone EBro03_SQ008_O11 5', mRNA sequence.
ACCESSION   BQ765891
VERSION     BQ765891.1  GI:21974365
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 658)
  AUTHORS   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
            Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
  TITLE     Development of Barley Transcriptome Resources
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .658
                     /organism="Hordeum vulgare"
                     /cultivar="Optic"
                     /db_xref="taxon:4513"
                     /clone="EBro03_SQ008_O11"
                     /clone_lib="root, 3 week, waterlogged, cv Optic, EBro03"
                     /tissue_type="root"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
                     Non-normalised library, directionally cloned into pSPORT1.
                     Derived from roots of 3 week old waterlogged barley
                     plants.  Developed as part of the barley transcriptome
                     resources of BBSRC/SEERAD funded cereal IGF (Investigating
                     Gene Function) project."
BASE COUNT      127 a    203 c    202 g    125 t      1 others
ORIGIN

  Query Match             32.5%;  Score 29.6;  DB 14;  Length 658;
  Best Local Similarity   59.5%;  Pred. No. 1.9e+02;
  Matches   50;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      6 GCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGAT 65
          |||    || | ||| |||||  ||   |||| || ||  | | || |   || | ||||
Db     99 GCGGCTGGGCACGGCTACCGCGTCCATGCGCCGAACCTAGCCGCGTCGGGCATCGATGAC 158

Qy     66 GGCCGGTCGACCGGGTTGACCGCG 89
          |||||  | | | | || |  ||
Db    159 CGCCGGCTGCCAGAGGTGGCAACG 182


RESULT 10
BM691860/c
LOCUS       BM691860                 488 bp    mRNA    linear   EST 28-FEB-2002
DEFINITION  UI-E-CL1-aet-h-03-0-UI.r1 UI-E-CL1 Homo sapiens cDNA clone
            UI-E-CL1-aet-h-03-0-UI 5', mRNA sequence.
ACCESSION   BM691860
VERSION     BM691860.1  GI:19005118
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 488)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
             cDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .488
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CL1-aet-h-03-0-UI"
                     /clone_lib="UI-E-CL1"
                     /tissue_type="human retina"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CL1 is a normalized cDNA library containing the
                     following tissue(s): retina. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is CCGCG. This library was
                     created for the program, Gene Discovery in the Visual
                     System, supported by National Eye Institute (NEI)."
BASE COUNT       75 a    175 c    163 g     75 t
ORIGIN

  Query Match             32.3%;  Score 29.4;  DB 14;  Length 488;
  Best Local Similarity   66.7%;  Pred. No. 2e+02;
```

```
  Matches   42;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      18 GGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGATGGCCGGTCGACC 77
           | |||||||  | | ||||  |||||||    ||| | || || |||  |   || || ||
Db     193 GCCGACCGCGACCGCGCGCAGATCCGGCTTTCGGTCGTCGACGATGGCGCGGGGCCGCCC 134

Qy      78 GGG 80
           | |
Db     133 GTG 131


RESULT 11
BH807613
LOCUS       BH807613                623 bp    DNA     linear   GSS 25-APR-2002
DEFINITION  1008072D06.x1 1008 - RescueMu Grid I Zea mays genomic, DNA
            sequence.
ACCESSION   BH807613
VERSION     BH807613.1  GI:20327180
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Walbot,V.
  TITLE     Maize genomic sequences found using engineered RescueMu transposon
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Possible ligation site so sequence was trimmed. Post-ligation
            sequence submitted separately.
            Plate: 1008072  row: 19
            Class: transposon-tagged.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Zea mays"
                     /cultivar="mixed background W23/A188/B73"
                     /db_xref="taxon:4577"
                     /clone_lib="1008 - RescueMu Grid I"
                     /tissue_type="leaf"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: leaf; Vector: RescueMu (engineered from
                     pBlueScript backbone); Site_1: BamHI; Site_2: BglII;
                     RescueMu is a 4.9 kb, modified maize Mu transposon
                     designed to allow plasmid rescue from total genomic DNA.
                     Mu elements insert preferentially into transcription
                     units. For more information on RescueMu, go to the  web
                     site www.zmdb.iastate.edu and follow the links for
                     'RescueMu.' Grid I was grown at Berkeley in 2001. DNA was
                     extracted from leaf punches,  double digested using BamHI
                     and BglII, and ligated to form circular plasmids.  DH10B
                     cells were transformed and then screened on LB plates with
                     ampicillin."
BASE COUNT      129 a    165 c    219 g    110 t
ORIGIN

  Query Match           32.3%;  Score 29.4;  DB 17;  Length 623;
  Best Local Similarity   60.8%;  Pred. No. 2.1e+02;
  Matches   48;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      13 GGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGATGGCCGGT 72
           | |  |||||| || | | | | | | |  || |||||||  |  ||||  ||      ||
Db     400 GTGGCGGCGACGGCTTGTCGTCCCGGGTGTGGTGCTCGGCGAGGAACCTGGAGAATACGT 459

Qy      73 CGACCGGGTTGACCGCGCC 91
            | | ||| || | |||||
Db     460 TGTCGGGGCTGGCGGCGCC 478


RESULT 12
BB739356/c
LOCUS       BB739356                398 bp    mRNA    linear   EST 15-OCT-2001
DEFINITION  BB739356 RIKEN full-length enriched, 6 days neonate spleen Mus
            musculus cDNA clone F430113C10 3', mRNA sequence.
ACCESSION   BB739356
VERSION     BB739356.1  GI:16142361
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 398)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .398
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="F430113C10"
                     /clone_lib="RIKEN full-length enriched, 6 days neonate
                     spleen"
                     /tissue_type="spleen"
                     /dev_stage="6 days neonate"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
```

```
                source libraries are cloned unidirectionally with Oligo(dT
                )-Not primers.  References include: (1) Genome-wide
                expression profiling of mid-gestation  placenta and embryo
                using a 15,000 mouse developmental cDNA micrcarray,  2000,
                Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                Large-scale cDNA  analysis reveals phased gene expression
                patterns during preimplantation  mouse develclpment, 2000,
                Development, 127: 1737-1749; (3) Genome-wide  mapping of
                unselected transcripts from extraembryonic tissue of
                7.5-day  mouse embryos reveals enrichment in the t-complex
                and under-representation  on the X chromosome, 1998, Hum
                Mol Genet 7: 1967-1978."
BASE COUNT       62 a    144 c    115 g     77 t
ORIGIN

  Query Match             31.9%;  Score 29;  DB 10;  Length 398;
  Best Local Similarity   58.8%;  Pred. No. 2.4e+02;
  Matches   50;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy       3 GACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
           |||||| ||||||  ||  || |  |||||| || |   |  |     |   | | |
Db     172 GACGCGGAACGGGGAGGGCACAGAGTCTGGGTGCAGCAGCCCGTTGAAGTAGACCGACGT 113

Qy      63 GATGGCCGGTCGACCGGGTTGACCG 87
             |||||||||| |||| ||  |||
Db     112 AGAGGCCGGTCGCCCGGCTTTGCCG 88


RESULT 13
AV618984
LOCUS       AV618984                 523 bp    mRNA    linear   EST 15-DEC-2000
DEFINITION  AV618984 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
            reinhardtii cDNA clone LC002c01_r 5', mRNA sequence.
ACCESSION   AV618984
VERSION     AV618984.1  GI:10768159
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
            Nakamura,Y. and Tabata,S.
  TITLE     Generation of expressed sequence tags from low-CO2 and high-CO2
            adapted cells of Chlamydomonas reinhardtii
  JOURNAL   DNA Res. 7 (5), 305-307 (2000)
  MEDLINE   20539644
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .523
                     /organism="Chlamydomonas reinhardtii"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LC002c01_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT       82 a    118 c    212 g    111 t
ORIGIN

  Query Match             31.9%;  Score 29;  DB 10;  Length 523;
  Best Local Similarity   61.0%;  Pred. No. 2.6e+02;
  Matches   47;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       8 GCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATGATGG 67
           ||| ||| ||| | | |  |  ||  ||||||| ||||  | ||||   |  |||||  |

Db     446 GCATCGGCATGCCAAGCAGAAGTGCTCGCTGATGCGGGCGTTGGCCGCGACACATGAGCG 505

Qy      68 CCGGTCGACCGGGTTGA 84
           | | | |  || |  ||
Db     506 CTGCTGGGGCGCGGGGA 522


RESULT 14
BE219908/c
LOCUS       BE219908                 549 bp    mRNA    linear   EST 03-JUL-2000
DEFINITION  hv64d03.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:3178181 3',
            mRNA sequence.
ACCESSION   BE219908
VERSION     BE219908.1  GI:8907226
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 485.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3178181"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo.  "
BASE COUNT      149 a    133 c    149 g    117 t      1 others
ORIGIN

  Query Match             31.9%;  Score 29;  DB 10;  Length 549;
  Best Local Similarity   61.0%;  Pred. No. 2.7e+02;
  Matches   47;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           || ||||    | || |||| | ||  | ||||  ||  || | ||   ||  ||  ||
Db      78 CGCCGCGGGCGGCGACGGCGGCGGCGGCCGGGCCGTGCGCCTGCGCGGCGCGCTCGGCCT 19

Qy      62 TGATGGCCGGTCGACCG 78
           |||||| ||  || |||
Db      18 TGATGGGCGCGCGGCCG 2


RESULT 15
AI920329
LOCUS       AI920329                 224 bp    mRNA    linear   EST 29-JUL-1999
```

```
DEFINITION  603019C06.x1 603 - stressed root cDNA library from Wang/Bohnert lab
            Zea mays cDNA, mRNA sequence.
ACCESSION   AI920329
VERSION     AI920329.1  GI:5649969
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta, Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 224)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 603019  row: C  column: 06.
FEATURES             Location/Qualifiers
     source          1. .224
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="603 - stressed root cDNA library from
                     Wang/Bohnert lab"
                     /tissue_type="seedling"
                     /dev_stage="salt stress"
                     /lab_host="E. coli XL Gold"
                     /note="Organ: root; Vector: pBluescriptII SK(+) XR;
                     Seedling stressed root cDNA library from Wang/Bohnert lab"
BASE COUNT       38 a    102 c     51 g     33 t
ORIGIN

  Query Match             31.4%;  Score 28.6;  DB 9;  Length 224;
  Best Local Similarity   57.1%;  Pred. No. 2.7e+02;
  Matches   52;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||  | | |  | |   |||  |  ||| | | ||  || || | || || ||     ||
Db      25 CCCGCTCACTTCTGAGCGGCACCGTCATGTTGACGGGGAGCCCGTGCGCGTCCCAATCCC 84

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
             |||| |  |||| ||| ||   |||||||
Db      85 CCGATGACTTGTCGTCCGCGTGCGCCGCGCC 115


Search completed: November  5, 2002, 16:17:28
Job time : 1009.2 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 25.8918 Seconds
(without alignments)
1077.852 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065
Perfect score: 91
Sequence: 1 ccgacgcgcaacgggatggc.....tcgaccgggttgaccgcgcc 91

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	91	100.0	68750	3	US-09-335-409-1	Sequence 1, Appli
2	91	100.0	68750	4	US-09-568-102-1	Sequence 1, Appli
3	91	100.0	68750	4	US-09-567-969-1	Sequence 1, Appli
4	91	100.0	68750	4	US-09-568-480-1	Sequence 1, Appli
5	91	100.0	68750	4	US-09-568-486-1	Sequence 1, Appli
6	91	100.0	68750	4	US-09-568-472-1	Sequence 1, Appli
7	91	100.0	68750	4	US-09-567-899-1	Sequence 1, Appli
8	91	100.0	71989	4	US-09-443-501A-2	Sequence 2, Appli
c 9	45.2	49.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 10	45.2	49.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
11	44.6	49.0	33529	4	US-09-144-085-3	Sequence 3, Appli
12	44.6	49.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
13	44.6	49.0	80161	3	US-09-036-987A-1	Sequence 1, Appli
14	44.6	49.0	80161	4	US-09-370-700-1	Sequence 1, Appli
15	43	47.3	643	4	US-08-861-774E-53	Sequence 53, Appl
16	43	47.3	20235	1	US-07-642-734C-3	Sequence 3, Appli
17	43	47.3	20235	3	US-08-439-009A-3	Sequence 3, Appli
c 18	42	46.2	507	4	US-09-060-756-680	Sequence 680, App
19	42	46.2	28958	1	US-08-258-261B-6	Sequence 6, Appli
20	42	46.2	28958	1	US-08-456-837-6	Sequence 6, Appli
21	42	46.2	28958	1	US-08-457-342-6	Sequence 6, Appli
22	42	46.2	28958	1	US-08-457-646A-6	Sequence 6, Appli
23	42	46.2	28958	1	US-08-458-076A-6	Sequence 6, Appli
24	42	46.2	28958	1	US-08-764-233A-4	Sequence 4, Appli
25	42	46.2	28958	1	US-08-457-335A-6	Sequence 6, Appli
26	42	46.2	28958	1	US-08-729-214-6	Sequence 6, Appli
27	42	46.2	28958	3	US-09-028-934-6	Sequence 6, Appli
28	42	46.2	49377	1	US-08-764-233A-1	Sequence 1, Appli
29	41.4	45.5	1462	4	US-09-434-288-4	Sequence 4, Appli
30	41.4	45.5	11220	4	US-09-105-537-32	Sequence 32, Appl
31	41.4	45.5	15872	4	US-09-105-537-1	Sequence 1, Appli
32	41.4	45.5	36778	4	US-09-105-537-5	Sequence 5, Appli
33	41.4	45.5	38506	3	US-09-320-878-19	Sequence 19, Appl
34	41.4	45.5	50937	4	US-09-428-517-1	Sequence 1, Appli
35	39.8	43.7	751	3	US-09-010-809-4	Sequence 4, Appli
36	39.8	43.7	4041	4	US-09-105-537-36	Sequence 36, Appl
37	39.8	43.7	44377	2	US-08-804-227C-7	Sequence 7, Appli
38	39.8	43.7	44377	2	US-08-804-198-1	Sequence 1, Appli
39	38.2	42.0	423	4	US-09-144-085-5	Sequence 5, Appli
40	38.2	42.0	754	4	US-08-861-774E-47	Sequence 47, Appl
41	38.2	42.0	13842	4	US-09-105-537-30	Sequence 30, Appl
42	36.8	40.4	643	4	US-08-861-774E-31	Sequence 31, Appl
43	36.6	40.2	13987	2	US-08-804-227C-13	Sequence 13, Appl
44	35	38.5	8035	4	US-09-215-694-29	Sequence 29, Appl
45	35	38.5	31328	4	US-09-215-694-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-09-335-409-1
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 100.0%; Score 91; DB 3; Length 68750;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677
```

RESULT 2
US-09-568-102-1
; Sequence 1, Application US/09568102
; Patent No. 6346404
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,102

```
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-102-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 3
US-09-567-969-1
; Sequence 1, Application US/09567969
; Patent No. 6355457
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,969
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-969-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 4
US-09-568-480-1
; Sequence 1, Application US/09568480
; Patent No. 6355458
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,480
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-480-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 5
US-09-568-486-1
; Sequence 1, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,486
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-486-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 6
US-09-568-472-1
; Sequence 1, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
```

```
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/568,472
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-568-472-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 7
US-09-567-899-1
; Sequence 1, Application US/09567899
; Patent No. 6383787
; GENERAL INFORMATION:
;  APPLICANT: Schupp, Thomas
;  APPLICANT:  Ligon, James
;  APPLICANT:  Molnar, Istvan
;  APPLICANT:  Zirkle, Ross
;  APPLICANT:  Cyr, Devon
;  APPLICANT:  Goerlach, Joern
;  TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
;  FILE REFERENCE: 4-30582A
;  CURRENT APPLICATION NUMBER: US/09/567,899
;  CURRENT FILING DATE:  2000-05-10
;  PRIOR APPLICATION NUMBER: 09/335,409
;  PRIOR FILING DATE: 1999-06-17
;  NUMBER OF SEQ ID NOS: 30
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 68750
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-567-899-1

  Query Match             100.0%;  Score 91;  DB 4;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 8
US-09-443-501A-2
; Sequence 2, Application US/09443501A
; Patent No. 6303342
; GENERAL INFORMATION:
;  APPLICANT: Kosan Biosciences, Inc.
;  APPLICANT:  Julien, Bryan
;  APPLICANT:  Katz, Leonard
;  APPLICANT:  Khosla, Chaitan
;  APPLICANT:  Tang, Li
;  APPLICANT:  Ziermann, Rainer
;  TITLE OF INVENTION: Recombinant Methods and Materials for Producing
;  TITLE OF INVENTION:  Epothilone and Epothilone Derivatives
;  FILE REFERENCE: 30062-20031.00
;  CURRENT APPLICATION NUMBER: US/09/443,501A
;  CURRENT FILING DATE:  1999-11-19
;  PRIOR APPLICATION NUMBER: US 60/130,560
;  PRIOR FILING DATE: 1999-04-22
;  PRIOR APPLICATION NUMBER: US 60/122,620
;  PRIOR FILING DATE: 1999-03-03
;  PRIOR APPLICATION NUMBER: US 60/119,386
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: US 60/109,401
;  PRIOR FILING DATE: 1998-11-20
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 71989
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Synthetic construct
US-09-443-501A-2

  Query Match             100.0%;  Score 91;  DB 4;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16975 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 17034

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  17035 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 17065


RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             49.7%;  Score 45.2;  DB 4;  Length 4403765;
  Best Local Similarity   68.9%;  Pred. No. 6.4e-05;
  Matches   62;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy      2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
```

```
          |||||| |  || || ||||||   ||| |||  |||||   |||   ||  |||||||
Db 4252308 CGACGCCCGCCGCGACGGCGACGCGATCCTGGCCGTGATCGCCGGCAGCGCGGTCAACCA 4252249

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           || |||||||| | |||  ||| |||  |
Db 4252248 CGACGGCCGGTCCAACGGCCTGATCGCACC 4252219


RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             49.7%;  Score 45.2;  DB 4;  Length 4411529;
  Best Local Similarity   68.9%;  Pred. No. 6.4e-05;
  Matches   62;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           |||||| |  || || ||||||   ||| |||  |||||   |||   ||  |||||||
Db 4260052 CGACGCCCGCCGCGACGGCGACGCGATCCTGGCCGTGATCGCCGGCAGCGCGGTCAACCA 4259993

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           || |||||||| | |||  ||| ||| ||
Db 4259992 CGACGGCCGGTCCAACGGCCTGATCGCACC 4259963


RESULT 11
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             49.0%;  Score 44.6;  DB 4;  Length 33529;
  Best Local Similarity   68.1%;  Pred. No. 7.3e-05;
  Matches   62;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | |||||||| || || ||||||||  |   || | ||||||| ||||| ||  | ||||
Db  21422 CTGACGCGCAGCGCGACGGCGACCGTGTGCTGGGGGTGATCCGTGGCTCTGCGGTGAACC 21481

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || || ||       ||  |||| |||||
Db  21482 AGGACGGTCGCAGCCAGGGTCTGACGGCGCC 21512


RESULT 12
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             49.0%;  Score 44.6;  DB 2;  Length 43280;
  Best Local Similarity   68.1%;  Pred. No. 7.4e-05;
  Matches   62;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | |||||  | ||  | ||| | |   |   |||| || |||| ||||||||  ||||||
Db  25420 CGGACGCCGAGCGCAACGGCCATCCGGTGCTGGCGGTGGTCCGCGGCTCGGCGGTCAACC 25479
```

```
Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || |||   || | |||| |||| |||||
Db   25480 AGGACGGCGCCTCCAACGGGCTGACGGCGCC 25510


RESULT 13
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             49.0%;  Score 44.6;  DB 3;  Length 80161;
  Best Local Similarity   68.1%;  Pred. No. 7.7e-05;
  Matches   62;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | || || |  ||| ||||  | ||  | | |||| || | || || |||||  | || |
Db   71219 CGGATGCTCGGCGGAATGGGCATCGGGTGTTGGCGGTGGTTCGTGGGTCGGCGGTGAATC 71278

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |||||   |||||  || ||||||||||||
Db   71279 AGGATGGTGCGTCGAATGGCTTGACCGCGCC 71309


RESULT 14
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             49.0%;  Score 44.6;  DB 4;  Length 80161;
  Best Local Similarity   68.1%;  Pred. No. 7.7e-05;
  Matches   62;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | || || |  ||| ||||  | ||  | | |||| || | || || |||||  | || |
Db   71219 CGGATGCTCGGCGGAATGGGCATCGGGTGTTGGCGGTGGTTCGTGGGTCGGCGGTGAATC 71278

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |||||   |||||  || ||||||||||||
Db   71279 AGGATGGTGCGTCGAATGGCTTGACCGCGCC 71309


RESULT 15
US-08-861-774E-53
; Sequence 53, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/08/861,774E
;  CURRENT FILING DATE:  1997-05-22
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-08-861-774E-53

  Query Match             47.3%;  Score 43;  DB 4;  Length 643;
  Best Local Similarity   67.0%;  Pred. No. 0.00017;
  Matches   61;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           || |||| ||||| ||||| |||   || | |||  | ||    || |||||  | ||||
Db     471 CCAACGCTCAACGAGATGGAGACAATATTTTGGCAGTTATTAAAGGTTCGGCAGTTAACC 530

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | ||||| ||    | ||| ||||| || ||
Db     531 AAGATGGTCGCAGCAACGGATTGACAGCACC 561


Search completed: November  5, 2002, 13:04:40
Job time : 1727.89 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 14:00:35 ; Search time 73.2088 Seconds
(without alignments)
414.118 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065
Perfect score: 91
Sequence: 1 ccgacgcgcaacgggatggc.....tcgaccgggttgaccgcgcc 91

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 310279 seqs, 166577418 residues

Total number of hits satisfying chosen parameters: 620558

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	43	47.3	643	10	US-09-924-256A-53	Sequence 53, Appl
2	41.4	45.5	11220	10	US-09-861-289-32	Sequence 32, Appl
3	41.4	45.5	15872	10	US-09-861-289-1	Sequence 1, Appli
4	41.4	45.5	36778	10	US-09-861-289-5	Sequence 5, Appli
5	39.8	43.7	4041	10	US-09-861-289-36	Sequence 36, Appl
6	38.2	42.0	754	10	US-09-924-256A-47	Sequence 47, Appl
7	38.2	42.0	13842	10	US-09-861-289-30	Sequence 30, Appl
8	36.8	40.4	643	10	US-09-924-256A-31	Sequence 31, Appl
9	36.6	40.2	6000	10	US-09-775-938A-29	Sequence 29, Appl
10	34.6	38.0	649	10	US-09-924-256A-61	Sequence 61, Appl
11	33.4	36.7	4689	10	US-09-861-289-34	Sequence 34, Appl
12	32.4	35.6	655	10	US-09-924-256A-45	Sequence 45, Appl
13	32.4	35.6	703	10	US-09-924-256A-51	Sequence 51, Appl
14	31.8	34.9	637	10	US-09-924-256A-41	Sequence 41, Appl
15	31	34.1	658	10	US-09-924-256A-69	Sequence 69, Appl
16	29	31.9	722	10	US-09-924-256A-49	Sequence 49, Appl
17	28.2	31.0	655	10	US-09-924-256A-55	Sequence 55, Appl
c 18	27.2	29.9	888	10	US-09-815-242-7736	Sequence 7736, Ap
c 19	26.8	29.5	1254	10	US-09-815-242-4025	Sequence 4025, Ap
20	26.6	29.2	1329	10	US-09-815-242-7893	Sequence 7893, Ap
c 21	26.2	28.8	3955	10	US-09-855-722-4	Sequence 4, Appli
c 22	26	28.6	13029	10	US-09-815-242-4052	Sequence 4052, Ap
23	25.8	28.4	680	10	US-09-924-256A-37	Sequence 37, Appl
24	25.8	28.4	1409	10	US-09-895-943-1	Sequence 1, Appli
25	25.6	28.1	312	10	US-09-960-352-1294	Sequence 1294, Ap
26	25.4	27.9	761	10	US-09-924-256A-79	Sequence 79, Appl
27	25.2	27.7	1299	10	US-09-815-242-7603	Sequence 7603, Ap
c 28	25	27.5	262	10	US-09-878-574-7566	Sequence 7566, Ap
c 29	25	27.5	871	10	US-09-974-300-307	Sequence 307, App
30	25	27.5	3489	10	US-09-815-242-7731	Sequence 7731, Ap
c 31	25	27.5	62804	12	US-10-096-960-3	Sequence 3, Appli
c 32	24.8	27.3	1083	10	US-09-815-242-7913	Sequence 7913, Ap
c 33	24.8	27.3	3936	10	US-09-900-237-9	Sequence 9, Appli
34	24.8	27.3	17710	10	US-09-956-004-70	Sequence 70, Appl
c 35	24.6	27.0	250	10	US-09-923-876-2514	Sequence 2514, Ap
36	24.6	27.0	691	10	US-09-924-256A-39	Sequence 39, Appl
37	24.6	27.0	903	10	US-09-764-860-293	Sequence 293, App
38	24.6	27.0	2618	10	US-09-954-456-1632	Sequence 1632, Ap
c 39	24.6	27.0	4702	10	US-09-880-107-1545	Sequence 1545, Ap
c 40	24.6	27.0	9815	10	US-09-764-847-1601	Sequence 1601, Ap
41	24.6	27.0	11842	10	US-09-764-860-1187	Sequence 1187, Ap
42	24.4	26.8	637	10	US-09-924-256A-33	Sequence 33, Appl
43	24.4	26.8	643	10	US-09-924-256A-43	Sequence 43, Appl
44	24.4	26.8	765	10	US-09-808-387-31	Sequence 31, Appl
45	24.4	26.8	779	10	US-09-731-872-131	Sequence 131, App

ALIGNMENTS

RESULT 1
US-09-924-256A-53
; Sequence 53, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Peltigera neopolydactyla
US-09-924-256A-53

Query Match 47.3%; Score 43; DB 10; Length 643;
Best Local Similarity 67.0%; Pred. No. 3.3e-05;
Matches 61; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           || |||| ||||| ||||| |||   || | |||  | ||    || |||||  | ||||
Db     471 CCAACGCTCAACGAGATGGAGACAATATTTTGGCAGTTATTAAAGGTTCGGCAGTTAACC 530

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | ||||| ||    | ||| ||||| || ||
Db     531 AAGATGGTCGCAGCAACGGATTGACAGCACC 561
```

RESULT 2
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:

```
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             45.5%;  Score 41.4;  DB 10;  Length 11220;
  Best Local Similarity   65.9%;  Pred. No. 0.00013;
  Matches   60;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTCGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||  ||  | ||| | || |||   ||  || ||||||||| | ||| ||||||
Db    851 CCGACGCGCGTCGCAAGGGCCATCGGATCCTCGCCGTGGTCCGGGGCACCGCCGTCAACC 910

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||      | |||  | || || |
Db    911 AGGACGGCGCCAGCAGCGGCCTCACGGCTCC 941


RESULT 3
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             45.5%;  Score 41.4;  DB 10;  Length 15872;
  Best Local Similarity   65.9%;  Pred. No. 0.00014;
  Matches   60;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTCGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          | |||||  | ||  | ||| ||| | |    ||  ||||||||||   ||  ||||||
Db   3843 CGGACGCCGAGCGCAACGGCCACCCCGTGCTCGCCGTGATCCGGGGCAGCGCGGTCAACC 3902

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||   || | |||| | ||||| ||
Db   3903 AGGACGGCGCCTCCAACGGGCTCACCGCCCC 3933


RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             45.5%;  Score 41.4;  DB 10;  Length 36778;
  Best Local Similarity   65.9%;  Pred. No. 0.00014;
  Matches   60;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||  ||  | ||| | || |||   ||  || ||||||||| | ||| ||||||
Db  16538 CCGACGCGCGTCGCAAGGGCCATCGGATCCTCGCCGTGGTCCGGGGCACCGCCGTCAACC 16597

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||      | |||  | || || ||
Db  16598 AGGACGGCGCCAGCAGCGGCCTCACGGCTCC 16628


RESULT 5
US-09-861-289-36
; Sequence 36, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 36
;   LENGTH: 4041
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-36

  Query Match             43.7%;  Score 39.8;  DB 10;  Length 4041;
  Best Local Similarity   64.8%;  Pred. No. 0.00039;
  Matches   59;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||  ||  | ||| |||| ||  |||  |  | || |||   ||||||||||
Db    848 CCGACGCGCGGCGCAAGGGCCACCGGGTCCTGGCCGTCGTGCGCGGCAGCGCCATCAACC 907

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||  |   | |||  | || || ||
Db    908 AGGACGGCGCGAGCAACGGCCTCACGGCTCC 938


RESULT 6
US-09-924-256A-47
; Sequence 47, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
```

```
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 754
;   TYPE: DNA
;   ORGANISM: Parmelia sulcata
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (753)
;   OTHER INFORMATION: n=a, c, g or t
US-09-924-256A-47

  Query Match            42.0%;  Score 38.2;  DB 10;  Length 754;
  Best Local Similarity   63.7%;  Pred. No. 0.0011;
  Matches   58;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          | ||||| | ||| |||  ||||    || | ||  |||| |||| |||  |    ||||
Db    581 CGGACGCTCTACGCGATAACGACACTGTCCGTGCAGTGATTCGGGCCTCTTCGTCTAACC 640

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | ||||||  |||  ||||  | ||   |||
Db    641 AGGATGGCAAGTCTCCCGGTATCACACAGCC 671


RESULT 7
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match            42.0%;  Score 38.2;  DB 10;  Length 13842;
  Best Local Similarity   63.7%;  Pred. No. 0.0013;
  Matches   58;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          | |||||||  ||  | ||  |||   |    ||  || |||| ||  | ||| ||||||
Db   8555 CGGACGCGCGGCGCAACGGTCACCCGGTGCTCGCCGTGGTCCGCGGTACCGCCGTCAACC 8614

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||      | |||| ||||||||||
Db   8615 AGGACGGCGCCAGCAACGGGCTGACCGCGCC 8645


RESULT 8
US-09-924-256A-31
; Sequence 31, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
;   LENGTH: 643
;   TYPE: DNA
;   ORGANISM: Leptogium corniculatum
US-09-924-256A-31

  Query Match            40.4%;  Score 36.8;  DB 10;  Length 643;
  Best Local Similarity   63.6%;  Pred. No. 0.0029;
  Matches   56;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      4 ACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCATG 63
          | ||| ||   |||||||| |  || |  |||   |||  ||| || |||   || ||||
Db    474 AGGCGGAAGCAGATGGCGATCTAATTTATGCGACAATCAAGGGGTCAGCCTCGAATCATG 533

Qy     64 ATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||| | |||  ||||  | ||||  ||
Db    534 GTGGACAGTCCGCCGGCCTCACCGTACC 561


RESULT 9
US-09-775-938A-29
; Sequence 29, Application US/09775938A
; Patent No. US20020081665A1
; GENERAL INFORMATION:
;  APPLICANT: Haygood, M.
;  APPLICANT:  Davidson, S.K.
;  APPLICANT:  Allen, S.W.
;  APPLICANT:  Hildebrand, M.
;  TITLE OF INVENTION: Bryostatins, Bryopyrans and Polyketides: Compositions and Meth
;  FILE REFERENCE: 1133.010US1
;  CURRENT APPLICATION NUMBER: US/09/775,938A
;  CURRENT FILING DATE:  2001-08-30
;  PRIOR APPLICATION NUMBER: PCT/US00/21326
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/147,283
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29
;   LENGTH: 6000
;   TYPE: DNA
;   ORGANISM: Endobugula sertula
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (386)...(388)
;   OTHER INFORMATION: TAG may represent a transposase open reading
;   OTHER INFORMATION: frame.
;   NAME/KEY: misc_feature
;   LOCATION: (444)...(449)
;   OTHER INFORMATION: TTGAAA may be a possible -35 trascription control
;   OTHER INFORMATION: sequence.
;   NAME/KEY: misc_feature
;   LOCATION: (458)...(463)
;   OTHER INFORMATION: GATAAT may be a possible -10 trascription control
;   OTHER INFORMATION: sequence.
;   NAME/KEY: misc_feature
;   LOCATION: (474)...(502)
;   OTHER INFORMATION: ATCAATAAAAA and TTTTTATTGAT are inverted repeats.
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (576)...(583)
;   OTHER INFORMATION: TGAGGAAT may be a possible SD sequence.
;   NAME/KEY: misc_feature
;   LOCATION: (565)...(567)
;   OTHER INFORMATION: ATG encoding M is presumptive start of PKS Open
;   OTHER INFORMATION: reading frame.
;   NAME/KEY: misc_feature
;   LOCATION: (589)...(591)
;   OTHER INFORMATION: GTG encoding V is is possible alternative start of
;   OTHER INFORMATION: PKS Open reading frame.
US-09-775-938A-29

  Query Match             40.2%;  Score 36.6;  DB 10;  Length 6000;
  Best Local Similarity   62.6%;  Pred. No. 0.004;
  Matches   57;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | || |||   || ||||| || | |||  | ||  |||| ||||||| ||   | || |
Db    2627 CGGATGCGGTGCGTGATGGTGATCCCATTCGTGCAGTGATACGGGGCTGGGGTGTGAATC 2686

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |||||  |    |  ||  | || |||||
Db    2687 AGGATGGTAGAAGTAATGGTATTACGGCGCC 2717


RESULT 10
US-09-924-256A-61
; Sequence 61, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 649
;   TYPE: DNA
;   ORGANISM: Pseudocyphellaria anthraspis
US-09-924-256A-61

  Query Match             38.0%;  Score 34.6;  DB 10;  Length 649;
  Best Local Similarity   61.8%;  Pred. No. 0.014;
  Matches   55;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy       3 GACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
           ||  ||    || ||||| ||| | ||  ||||  | ||||| ||| | |   |||||
Db     479 GATGCTATTCGAGATGGAGACTGTATAAGGGCAGTAATCCGCGGCACTGGGGTCAACTCC 538

Qy      63 GATGGCCGGTCGACCGGGTTGACCGCGCC 91
           ||||||  | |  | ||  | |||  |||
Db     539 GATGGCAAGACCCCTGGCATTACCTTGCC 567


RESULT 11
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             36.7%;  Score 33.4;  DB 10;  Length 4689;
  Best Local Similarity   60.4%;  Pred. No. 0.039;
  Matches   55;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           | ||||| |  ||  | ||  |||| |||   ||| |  |||| |||   ||  ||||||
Db     854 CGGACGCCCGCCGCAACGGACACCGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACC 913

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || |||       | |||  | || || ||
Db     914 AGGACGGCGCCAGCAACGGCCTCACGGCTCC 944


RESULT 12
US-09-924-256A-45
; Sequence 45, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 655
;   TYPE: DNA
;   ORGANISM: Parmelia sulcata
US-09-924-256A-45

  Query Match             35.6%;  Score 32.4;  DB 10;  Length 655;
  Best Local Similarity   60.0%;  Pred. No. 0.069;
  Matches   54;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           || |||||    ||| || |||| |||       || || ||||   |||   |||||||
Db     472 CGCCGCGCTGGCGGACGGAGACCACATTGAATGTCTCATACGGGAGACGGGACTCAACCA 531

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGCC 91
            || ||   | |  ||||  | |||  |||
Db     532 GGACGGTGCGACAGCCGGTCTCACCATGCC 561


RESULT 13
US-09-924-256A-51
; Sequence 51, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
```

```
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 703
;   TYPE: DNA
;   ORGANISM: Peltigera neopolydactyla
US-09-924-256A-51

  Query Match             35.6%;  Score 32.4;  DB 10;  Length 703;
  Best Local Similarity   60.0%;  Pred. No. 0.069;
  Matches   54;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       2 CGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCA 61
           ||  ||    || ||||| || |  |||  |    | |||||||   | |  || |||||
Db     472 CGCTGCAATTCGTGATGGAGATCATATCGAGTGCATTATCCGGGAGACAGGAATTAACCA 531

Qy      62 TGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           |||||| |  |  ||||  | |||  |||
Db     532 GGATGGCAGAACAGCCGGAATTACCATGCC 561


RESULT 14
US-09-924-256A-41
; Sequence 41, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Leptogium corniculatum
US-09-924-256A-41

  Query Match             34.9%;  Score 31.8;  DB 10;. Length 637;
  Best Local Similarity   59.3%;  Pred. No. 0.11;
  Matches   54;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| ||  ||   || || ||    ||||| |   |||| || |||| ||   | || |
Db     465 CCGATGCCGAAAAAGACGGTGATAATATCTGTGGTGTGATTCGAGGCTGGGGGGTGAATC 524

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || |||  | | |  ||  | || || ||
Db     525 AAGACGGCAAGACCAGTGGAATTACAGCACC 555


RESULT 15
US-09-924-256A-69
; Sequence 69, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
;  APPLICANT: Waters, Barbara
;  APPLICANT:  Miao, Vivian
;  APPLICANT:  Ho, Yap
;  APPLICANT:  Tong, Seow
;  TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
;  TITLE OF INVENTION:  BIOACTIVE MOLECULES
;  FILE REFERENCE: 9993-006
;  CURRENT APPLICATION NUMBER: US/09/924,256A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: 08/861,774
;  PRIOR FILING DATE: 2001-04-13
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
;   LENGTH: 658
;   TYPE: DNA
;   ORGANISM: Usnea florida
US-09-924-256A-69

  Query Match             34.1%;  Score 31;  DB 10;  Length 658;
  Best Local Similarity   58.4%;  Pred. No. 0.19;
  Matches   52;  Conservative    1;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       3 GACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
           |||||      |||||| :| ||||||  |    | ||||| |   | |    ||||||
Db     473 GACGCTATTGCGGATGGASATCGCATCGAGTGCATCATCCGTGAGACAGGGTCCAACCAA 532

Qy      63 GATGGCCGGTCGACCGGGTTGACCGCGCC 91
           || ||||  || |  ||  | || | |||
Db     533 GACGGCCATTCAAATGGTATCACGGTGCC 561


Search completed: November  5, 2002, 22:59:08
Job time : 88.2088 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 123.63 Seconds
(without alignments)
1657.625 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065
Perfect score: 91
Sequence: 1 ccgacgcgcaacgggatggc.....tcgaccgggttgaccgcgcc 91

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	91	100.0	68750	21	AAZ55887	Sorangium cellulos
	2	91	100.0	71989	21	AAA29349	Sorangium cellulos
	3	51	56.0	27541	22	AAD17185	Streptomyces nours
	4	51	56.0	125401	22	AAD17186	Streptomyces nours
	5	47.8	52.5	671	22	AAF89969	Partial nucleotide
	6	47.8	52.5	764	22	AAF89974	Partial nucleotide
	7	47.8	52.5	4306	22	AAF90039	Nucleotide sequenc
	8	47.8	52.5	6462	22	AAF90037	Nucleotide sequenc
	9	47.8	52.5	34071	22	AAF90033	Nucleotide sequenc
c	10	47.8	52.5	42717	22	AAF90032	Nucleotide sequenc
	11	47.8	52.5	77536	21	AAA14651	Nucleotide sequenc
	12	47.8	52.5	109519	22	AAS08693	Micromonospora DNA
	13	46.4	51.0	330	24	ABN76829	Human synthase-lik
	14	46.2	50.8	758	22	AAF89970	Partial nucleotide
	15	46.2	50.8	5292	22	AAF90036	Nucleotide sequenc
	16	46.2	50.8	65140	22	AAD17184	Streptomyces nours
	17	45.2	49.7	5544	22	AAF24891	Pimaricin biosynth
c	18	45.2	49.7	4403765	22	AAI99683	Mycobacterium tube
c	19	45.2	49.7	4411529	22	AAI99682	Mycobacterium tube
	20	44.6	49.0	4466	21	AAA14663	Nucleotide sequenc
	21	44.6	49.0	4478	21	AAA14661	Nucleotide sequenc
	22	44.6	49.0	4547	21	AAA14664	Nucleotide sequenc
	23	44.6	49.0	4571	21	AAA14662	Nucleotide sequenc
	24	44.6	49.0	4614	22	AAF31636	Mycobacterium tube
	25	44.6	49.0	5088	22	AAF90038	Nucleotide sequenc
	26	44.6	49.0	5628	22	AAF31635	Mycobacterium tube
	27	44.6	49.0	16767	22	AAF88339	S. spinosa DNA fra
	28	44.6	49.0	29736	22	AAF88317	S. spinosa DNA fra
	29	44.6	49.0	33529	23	AAS17367	DNA sequence of S.
	30	44.6	49.0	43280	18	AAT80413	Tylactone synthase
	31	44.6	49.0	50000	22	AAF88313	S. spinosa DNA fra
c	32	44.6	49.0	77536	21	AAA14651	Nucleotide sequenc
	33	44.6	49.0	80161	20	AAZ21501	DNA fragment of Sa
	34	44.2	48.6	330	24	ABN20329	Human ORFX polynuc
	35	43	47.3	643	20	AAV69543	P. neopolydactyla
	36	43	47.3	5481	22	AAF31638	Mycobacterium tube
	37	43	47.3	5485	22	AAH52061	Mycobacterium tube
	38	43	47.3	29879	14	AAQ46806	eryA region of S.
	39	42.6	46.8	20394	22	AAF24892	Pimaricin biosynth
c	40	42	46.2	507	21	ABQ63181	Mycobacterium tube
	41	42	46.2	28598	17	AAT06769	Sorangium cellulos
	42	42	46.2	28958	18	AAT89956	Sorangium cellulos
	43	42	46.2	28958	21	AAA75299	DNA sequence of So
	44	42	46.2	49377	19	AAV05287	The soraphen biosy
	45	41.8	45.9	671	22	AAF89977	Partial nucleotide

ALIGNMENTS

RESULT 1
AAZ55887
ID AAZ55887 standard; DNA; 68750 BP.
XX
AC AAZ55887;
XX
DT 10-APR-2000 (first entry)
XX
DE Sorangium cellulosum 68.75 kb contig.
XX
KW Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KW anticancer; ds.
XX
OS Sorangium cellulosum.
XX
FH Key Location/Qualifiers
FT CDS 1..1826
FT /*tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAY58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT /*tag= b
FT /product= "Orf 2 protein (AAY58581)"
FT CDS 3415..5556
FT /*tag= c
FT /product= "Orf 3 protein (AAY58582)"
FT CDS complement (5612..5992)
FT /*tag= d
FT /product= "Orf 4 protein (AAY58583)"
FT CDS 6226..6675
FT /*tag= e

```
FT                   /product= "Orf 5 protein (AAY58584)"
FT   CDS             7610..11875
FT                   /*tag= f
FT                   /product= "Type I polyketide synthase, EPOS A
FT                   (AAY58573)"
FT   CDS             11872..116104
FT                   /*tag= g
FT                   /product= "Non-ribosomal peptide synthetase, EPOS P
FT                   (AAY58574)"
FT   CDS             16251..21749
FT                   /*tag= h
FT                   /product= "Type I polyketide synthase, EPOS B
FT                   (AAY58575)"
FT   CDS             21746..43519
FT                   /*tag= i
FT                   /product= "Type I polyketide synthase, EPOS C
FT                   (AAY58576)"
FT   CDS             43524..54920
FT                   /*tag= j
FT                   /product= "Type I polyketide synthase, EPOS D
FT                   (AAY58577)"
FT   CDS             54935..62254
FT                   /*tag= k
FT                   /product= "Type I polyketide synthase, EPOS E
FT                   (AAY58578)"
FT   CDS             62369..63628
FT                   /*tag= l
FT                   /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT                   (AAY58579)"
FT   CDS             63779..64333
FT                   /*tag= m
FT                   /product= "Orf 6 protein (AAY58585)"
FT   CDS             complement (63853. 64290)
FT                   /*tag= n
FT                   /product= "Orf 7 protein (AAY58586)"
FT   CDS             64363..64920
FT                   /*tag= o
FT                   /product= "Orf 8 protein (AAY58587)"
FT   CDS             complement (64287..64727)
FT                   /*tag= p
FT                   /product= "Orf 9 protein (AAY58588)"
FT   CDS             65063..65767
FT                   /*tag= q
FT                   /product= "Orf 10 protein"
FT   CDS             complement (65008..65874)
FT                   /*tag= r
FT                   /product= "Orf 11 protein (AAY58590)"
FT   CDS             complement (65871..66338)
FT                   /*tag= s
FT                   /product= "Orf 12 protein (AAY58591)"
FT   CDS             66667..67137
FT                   /*tag= t
FT                   /product= "Orf 13 protein (AAY58592)"
FT   CDS             67334..68251
FT                   /*tag= u
FT                   /product= "Orf 14 protein (AAY58593)"
FT   CDS             68346..68750
FT                   /partial
FT                   /*tag= v
FT                   /product= "Partial Orf 15 protein (AAY58594)"
FT                   /note= "No termination codon given in the specification"
XX
PN   WO9966028-A2.
XX
PD   23-DEC-1999.
XX
PF   16-JUN-1999;   99WO-EP04171.
XX
PR   18-JUN-1998;   98US-0099504.
PR   24-SEP-1998;   98US-0101631.
PR   05-FEB-1999;   99US-0118906.
XX
PA   (NOVS ) NOVARTIS AG.
PA   (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Schupp T,  Ligon JM,  Molnar I,  Zirkle R,  Goerlach J,  Cyr D;
XX
DR   WPI; 2000-097741/08.
DR   P-PSDB; AAY58573, AAY58574, AAY58575, AAY58576, AAY58577, AAY58578,
DR           AAY58579, AAY58580, AAY58581, AAY58582, AAY58583, AAY58584,
DR           AAY58585, AAY58586, AAY58587, AAY58588, AAY58590, AAY58591,
DR           AAY58592, AAY58593, AAY58594.
XX
PT   New isolated epothilone synthase genes, used for the recombinant
PT   production of epothilone for use in cancer therapy -
XX
PS   Claim 14; Page 87-104; 174pp; English.
XX
CC   This sequence represents a 68.75 kb contig from Sorangium cellulosum
CC   comprising 22 open reading frames (ORFs) and includes genes encoding
CC   proteins involved in the biosynthesis of epothilones. Epothilones A and
CC   B are 16-membered macrocyclic polyketides with an acylcysteine-derived
CC   starter unit; polyketides being synthesised from two-carbon building
CC   blocks, the beta-carbon of which always carries a keto group. Each round
CC   of two-carbon addition is carried out by a complex of enzymes known as
CC   the polyketide synthase in a manner similar to fatty acid biosynthesis.
CC   EPOS A (AAY58573) and EPOS P (AAY58574) are involved in formation of
CC   the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
CC   and EPOS E (AAY58575-Y58578) are involved in polyketide backbone
CC   formation. EPO F (AAY58579) is an epothilone macrolactone oxidase, and
CC   the proteins Orf 3 (AAY58582) and Orf14 (AAY58593) are thought to be
CC   involved in transport. Epothilones mimic the biological activity of
CC   taxol, and may be substituted for taxol in cancer chemotherapeutic
CC   compositions. Epothilones exhibit a much lower drop in potency against a
CC   multiply drug-resistant cell line compared with taxol, and are
CC   considerably less efficiently exported from such cells by the multidrug
CC   resistance protein (MDR, or P-glycoprotein). Despite the potential of
CC   epothilones as anticancer agents, they are problematical to produce on a
CC   large scale. Epothilones are too complex for industrial scale chemical
CC   synthesis, and Sorangium cellulosum is difficult to ferment, producing
CC   poor yields of epothilones. The nucleic acids of the invention may be
CC   used for the recombinant production of epothilones in a heterologous host
CC   that is more amenable to fermentation.
XX
SQ   Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

  Query Match             100.0%;  Score 91;  DB 21;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           |||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 2
AAA29349
ID   AAA29349 standard; DNA; 71989 BP.
XX
AC   AAA29349;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Sorangium cellulosum epothilone polyketide synthase operon genomic DNA.
XX
KW   Epothilone; polyketide synthase; epoA; epoB; epoC; epoD; epoE; epoF;
KW   epoL; epoK; P450 epoxidase; ORFA; ORFB; promoter; enhancer; anti-fungal;
KW   tubulin polymerization assay; anti-tumour; cytostatic; ds.
XX
OS   Sorangium cellulosum.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             3..992
FT                   /*tag= a
FT                   /label= ORF_A
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             989..1501
FT                   /*tag= b
FT                   /label= ORF_B
FT                   /product= transposase
FT                   /note= "not part of the PKS"
FT   CDS             1998..6263
FT                   /*tag= c
FT                   /label= epoA_gene
FT                   /note= "encodes the loading domain"
FT   misc_RNA        2031..3548
FT                   /*tag= d
FT                   /note= "encodes ketide synthase (KS-Y) of the loading
FT                   domain"
FT   misc_RNA        3621..4661
FT                   /*tag= e
FT                   /note= "encodes acyl transferase (AT) of the loading
FT                   domain"
FT   misc_RNA        4917..5810
FT                   /*tag= f
FT                   /note= "encodes enoyl reductase (ER) of the loading
FT                   domain, potentially involved in formation of the
FT                   thiazole moiety"
FT   misc_RNA        5856..6155
FT                   /*tag= g
FT                   /note= "encodes acyl carrier protein (ACP) of the loading
FT                   domain"
FT   CDS             6260..10493
FT                   /*tag= h
FT                   /label= epoB_gene
FT                   /note= "encodes module 1, the NRPS module"
FT   misc_RNA        2031..3548
FT                   /*tag= i
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        2031..3548
FT                   /*tag= j
FT                   /note= "encodes condensation domain C2 of the NRPS
FT                   module"
FT   misc_RNA        6861..6887
FT                   /*tag= k
FT                   /note= "encodes heterocyclization signature sequence"
FT   misc_RNA        6861..6887
FT                   /*tag= l
FT                   /note= "encodes condensation domain C4 of the NRPS
FT                   module"
FT   misc_RNA        7358..7366
FT                   /*tag= m
FT                   /note= "encodes condensation domain C7 (partial) of the
FT                   NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= n
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        7898..7921
FT                   /*tag= o
FT                   /note= "encodes adenylation domain A1 of the NRPS module"
FT   misc_RNA        8261..8308
FT                   /*tag= p
FT                   /note= "encodes adenylation domain A3 of the NRPS module"
FT   misc_RNA        8411..8422
FT                   /*tag= q
FT                   /note= "encodes adenylation domain A4 of the NRPS module"
FT   misc_RNA        8861..8905
FT                   /*tag= r
FT                   /note= "encodes adenylation domain A6 of the NRPS module"
FT   misc_RNA        8966..8983
FT                   /*tag= s
FT                   /note= "encodes adenylation domain A7 of the NRPS module"
FT   misc_RNA        9090..9179
FT                   /*tag= t
FT                   /note= "encodes adenylation domain A8 of the NRPS module"
FT   misc_RNA        9183..9992
FT                   /*tag= u
FT                   /note= "encodes oxidation region for forming thiazole"
FT   misc_RNA        10121..10138
FT                   /*tag= v
FT                   /note= "encodes adenylation domain A10 of the NRPS
FT                   module"
FT   misc_RNA        10261..10306
FT                   /*tag= w
FT                   /note= "encodes thiolation domain (PCP) of the NRPS
FT                   module"
FT   CDS             10639..16137
FT                   /*tag= x
FT                   /label= epoC_gene
FT                   /note= "encodes module 2"
FT   misc_RNA        10654..12033
FT                   /*tag= y
FT                   /note= "encodes KS2, the KS domain of module 2"
FT   misc_RNA        12250..13287
FT                   /*tag= z
FT                   /note= "encodes AT2, the AT domain of module 2"
FT   misc_RNA        13327..13899
FT                   /*tag= aa
FT                   /note= "encodes dehydratase (DH) 2, the DH domain of
FT                   module 2"
FT   misc_RNA        14962..15756
FT                   /*tag= ab
FT                   /note= "encodes ketoreductase (KR) 2, the KR domain of
FT                   module 2"
FT   misc_RNA        15763..16008
FT                   /*tag= ac
FT                   /note= "encodes ACP2, the ACP domain of module 2"
FT   CDS             16134..37907
FT                   /*tag= ad
FT                   /label= epoD_gene
FT                   /note= "encodes modules 3-6"
FT   misc_RNA        16425..17606
FT                   /*tag= ae
FT                   /note= "encodes KS3"
FT   misc_RNA        17817..18857
FT                   /*tag= af
FT                   /note= "encodes AT3"
FT   misc_RNA        19581..20396
FT                   /*tag= ag
FT                   /note= "encodes KR3"
FT   misc_RNA        20424..20642
FT                   /*tag= ah
FT                   /note= "encodes ACP3"
FT   misc_RNA        20706..22082
FT                   /*tag= ai
FT                   /note= "encodes KS4"
FT   misc_RNA        22296..23336
FT                   /*tag= aj
FT                   /note= "encodes AT4"
FT   misc_RNA        24069..24647
FT                   /*tag= ak
FT                   /note= "encodes KR4"
FT   misc_RNA        24867..25151
FT                   /*tag= al
FT                   /note= "encodes ACP4"
FT   misc_RNA        25203..26576
FT                   /*tag= am
FT                   /note= "encodes KS5"
FT   misc_RNA        26793..27883
FT                   /*tag= an
FT                   /note= "encodes AT5"
FT   misc_RNA        27966..28574
FT                   /*tag= ao
FT                   /note= "encodes DH5"
FT   misc_RNA        29433..30287
FT                   /*tag= ap
```

```
FT                   /note= "encodes ER5"
FT   misc_RNA        30321..30869
FT                   /*tag= aq
FT                   /note= "encodes KR5"
FT   misc_RNA        31077..31373
FT                   /*tag= ar
FT                   /note= "encodes ACP5"
FT   misc_RNA        31440..32807
FT                   /*tag= as
FT                   /note= "encodes KS6"
FT   misc_RNA        33018..34067
FT                   /*tag= at
FT                   /note= "encodes AT6"
FT   misc_RNA        34107..34676
FT                   /*tag= au
FT                   /note= "encodes DH6"
FT   misc_RNA        35760..36641
FT                   /*tag= av
FT                   /note= "encodes ER6"
FT   misc_RNA        36705..37256
FT                   /*tag= aw
FT                   /note= "encodes KR6"
FT   misc_RNA        37470..37769
FT                   /*tag= ax
FT                   /note= "encodes ACP6"
FT   CDS             37912..49308
FT                   /*tag= ay
FT                   /label= epoE_gene
FT                   /note= "encodes modules 7 and 8"
FT   misc_RNA        38014..39375
FT                   /*tag= az
FT                   /note= "encodes KS7"
FT   misc_RNA        39589..40626
FT                   /*tag= ba
FT                   /note= "encodes AT7"
FT   misc_RNA        41341..41922
FT                   /*tag= bb
FT                   /note= "encodes KR7"
FT   misc_RNA        42181..42423

  Query Match             100.0%;  Score 91;  DB 21;  Length 71989;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-18;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16975 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 17034

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  17035 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 17065


RESULT 3
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             56.0%;  Score 51;  DB 22;  Length 27541;
  Best Local Similarity   72.5%;  Pred. No. 2.1e-06;
  Matches   66;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||| ||  | ||| |||   ||  |||  |  |||| ||||| ||||||||||
Db  17145 CCGACGCGCAGCGCAACGGCCACCCGGTCCTGGCCGTCCTCCGCGGCTCCGCCATCAACC 17204

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||   || | |||  ||||||||||
Db  17205 AGGACGGCGCCTCCAACGGCCTGACCGCGCC 17235


RESULT 4
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
```

```
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match            56.0%;  Score 51;  DB 22;  Length 125401;
  Best Local Similarity   72.5%;  Pred. No. 2.4e-06;
  Matches   66;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||| ||  | ||| |||   ||  |||  |  |||| ||||| ||||||||||
Db  17145 CCGACGCGCAGCGCAACGGCCACCCGGTCCTGGCCGTCCTCCGCGGCTCCGCCATCAACC 17204

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||   || | |||  ||||||||||
Db  17205 AGGACGGCGCCTCCAACGGCCTGACCGCGCC 17235


RESULT 5
AAF89969
ID   AAF89969 standard; DNA; 671 BP.
XX
AC   AAF89969;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83961.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 221-222; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 671 BP; 108 A; 223 C; 227 G; 113 T; 0 other;
```

```
  Query Match           52.5%;  Score 47.8;  DB 22;  Length 671;
  Best Local Similarity  70.3%;  Pred. No. 1.4e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTCGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||| |||   || || ||||| || ||   |||||||| || || |||||| ||||||
Db    482 CCGATGCGACGCGTGACGGCGATCGTATTCTGGCGCTGATTCGCGGATCGGCCGTCAACC 541

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | |  |||||    | |||  | || |||||
Db    542 ACGGCGGCCGCAGCAACGGCCTCACGGCGCC 572


RESULT 6
AAF89974
ID   AAF89974 standard; DNA; 764 BP.
XX
AC   AAF89974;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83966.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 224; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 764 BP; 114 A; 272 C; 268 G; 110 T; 0 other;

  Query Match           52.5%;  Score 47.8;  DB 22;  Length 764;
  Best Local Similarity  70.3%;  Pred. No. 1.4e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          |||||||||    ||| ||||| |  |||   |||||||||| || || ||  |||| |
Db    575 CCGACGCGCTGGCGGACGGCGATCCGATCGTCGCGCTGATCCGCGGATCCGCGGTCAATC 634

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||||      ||| || ||||||||
Db    635 AGGACGGCCGCAGCGGCGGCTTCACCGCGCC 665


RESULT 7
AAF90039
ID   AAF90039 standard; DNA; 4306 BP.
XX
AC   AAF90039;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4301
FT                   /*tag=  a
FT                   /product= "type I polyketide synthase"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83976.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Claim 35; Page 311-312; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 4306 BP; 720 A; 1344 C; 1451 G; 791 T; 0 other;

  Query Match           52.5%;  Score 47.8;  DB 22;  Length 4306;
  Best Local Similarity  70.3%;  Pred. No. 1.7e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;
```

```
Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| |||   || || ||||| || ||   ||||||||| || || |||||| ||||||
Db     830 CCGATGCGACGCGTGACGGCGATCGTATTCTGGCGCTGATTCGCGGATCGGCCGTCAACC 889

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |  |||||    | |||  | || |||||
Db     890 ACGGCGGCCGCAGCAACGGCCTCACGGCGCC 920
```

RESULT 8
AAF90037
ID AAF90037 standard; DNA; 6462 BP.
XX
AC AAF90037;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of a type I polyketide synthase.
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic;
KW type I polyketide synthase; ss.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT CDS 1..6462
FT /*tag= a
FT /product= "type I polyketide synthase"
FT /transl_except= "(1..3, aa: Met)"
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
DR P-PSDB; AAB83974.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics -
XX
PS Claim 35; Page 307-309; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC represent open reading frames (ORFS) of the coding strand of cosmid
CC a26g1, and encode type I polyketide synthases.
XX
SQ Sequence 6462 BP; 1073 A; 2067 C; 2135 G; 1187 T; 0 other;

Query Match 52.5%; Score 47.8; DB 22; Length 6462;
Best Local Similarity 70.3%; Pred. No. 1.7e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| || |    ||  || |||||  | |||||  || | || ||||| ||  |||| |
Db     857 CCGATGCACTCGCGGCCGGAGACCGGGTGTGGGCCGTGGTGCGCGGCTCCGCGGTCAATC 916

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |||||||| ||| ||||| | ||||| ||
Db     917 AGGATGGCCGCTCGGCCGGGCTCACCGCTCC 947
```

RESULT 9
AAF90033
ID AAF90033 standard; DNA; 34071 BP.
XX
AC AAF90033;
XX
DT 06-AUG-2001 (first entry)
XX
DE Nucleotide sequence of cosmid a26g1 (coding strand).
XX
KW Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS Synthetic.
XX
PN WO200140497-A2.
XX
PD 07-JUN-2001.
XX
PF 27-NOV-2000; 2000WO-FR03311.
XX
PR 29-NOV-1999; 99FR-0015032.
PR 07-JUN-2000; 2000US-0209800.
XX
PA (AVET) AVENTIS PHARMA SA.
XX
PI Jeannin P, Pernodet J, Guerineau M, Simonet P, Courtois S;
PI Cappellano C, Francou F, Raynal A, Ball M, Sezonov G, Tuphile K;
PI Frostegard A;
XX
DR WPI; 2001-374849/39.
XX
PT Collection of nucleic acids from environmental samples, useful for
PT identifying e.g. genes encoding polyketide synthases and derived
PT antibiotics -
XX
PS Example 14; Page 289-300; 356pp; French.
XX
CC The specification describes a method for the preparation of a collection
CC of nucleic acids from organisms in a soil sample. The method comprises
CC milling a dried sample to produce microparticles; suspending these in
CC liquid buffer; extraction of nucleic acids from the microparticle;
CC passing nucleic acid-containing solution through a molecular sieve;
CC passing nucleic acid-enriched fractions through an anion exchange
CC chromatography material; and recovering fractions containing purified
CC nucleic acids. The nucleic acids are sources for sequences that encode
CC either operons involved in a metabolic pathway (specifically polyketide
CC synthesis) or polypeptides, particularly for production of therapeutic
CC or agricultural compounds, especially polyketide antibiotics. The present
CC sequence represents cosmid a26g1 coding strand), which encodes different
CC polyketide synthases.
XX
SQ Sequence 34071 BP; 5791 A; 10858 C; 11089 G; 6333 T; 0 other;

Query Match 52.5%; Score 47.8; DB 22; Length 34071;
Best Local Similarity 70.3%; Pred. No. 2e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```
Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| || |    ||  || |||||  | |||||  || | || ||||| ||  |||| |
Db  19080 CCGATGCACTCGCGGCCGGAGACCGGGTGTGGGCCGTGGTGCGCGGCTCCGCGGTCAATC 19139

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | |||||||| ||| ||||| | ||||| ||
```

```
Db  19140 AGGATGGCCGCTCGGCCGGGCTCACCGCTCC 19170


RESULT 10
AAF90032/c
ID   AAF90032 standard; DNA; 42717 BP.
XX
AC   AAF90032;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of cosmid a26g1 (non-coding strand).
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic; ss.
XX
OS   Synthetic.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics  -
XX
PS   Example 14; Page 274-288; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. The present
CC   sequence represents cosmid a26g1 (non-coding strand). The sense strand
CC   encodes different polyketide synthases.
XX
SQ   Sequence 42717 BP; 8230 A; 13520 C; 13184 G; 7782 T; 1 other;

  Query Match             52.5%;  Score 47.8;  DB 22;  Length 42717;
  Best Local Similarity   70.3%;  Pred. No. 2.1e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| |||   || || ||||| || ||   ||||||||| || || |||||| ||||||
Db    3525 CCGATGCGACGCGTGACGGCGATCGTATTCTGGCGCTGATTCGCGGATCGGCCGTCAACC 3466

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
             | |  |||||    | |||  | || |||||
Db    3465 ACGGCGGCCGCAGCAACGGCCTCACGGCGCC 3435


RESULT 11
AAA14651
ID   AAA14651 standard; DNA; 77536 BP.
XX
AC   AAA14651;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW   FK-506 binding protein; polyketide compound; transplant rejection;
KW   graft-versus-host disease; uveitis; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (412..1836)
FT                   /*tag= a
FT                   /note= "fkbW gene"
FT   CDS             complement (2020..3579)
FT                   /*tag= b
FT                   /note= "fkbV gene"
FT   CDS             3969..4496
FT                   /*tag= c
FT                   /note= "fkbR2 gene"
FT   CDS             complement (4595..5488)
FT                   /*tag= d
FT                   /note= "fkbR1 gene"
FT   CDS             5601..6818
FT                   /*tag= e
FT                   /note= "fkbE gene"
FT   CDS             6808..8052
FT                   /*tag= f
FT                   /note= "fkbF gene"
FT   CDS             8156..8824
FT                   /*tag= g
FT                   /note= "fkbG gene"
FT   CDS             complement (9122..9883)
FT                   /*tag= h
FT                   /note= "fkbH gene"
FT   CDS             complement (9894..10994)
FT                   /*tag= i
FT                   /note= "fkbI gene"
FT   CDS             complement (10987..11247)
FT                   /*tag= j
FT                   /note= "fkbJ gene"
FT   CDS             complement (11244..12092)
FT                   /*tag= k
FT                   /note= "fkbK gene"
FT   CDS             complement (12113..13150)
FT                   /*tag= l
FT                   /note= "fkbL gene"
FT   CDS             complement (13212..23988)
FT                   /*tag= m
FT                   /note= "fkbC gene"
FT   misc_feature    complement (13452..13662)
FT                   /*tag= n
FT                   /note= "ACP6"
FT   misc_feature    complement (13761..14394)
FT                   /*tag= o
FT                   /note= "KR6"
FT   misc_feature    complement (14517..15294)
FT                   /*tag= p
FT                   /note= "ER6"
FT   misc_feature    complement (15438..16587)
FT                   /*tag= q
FT                   /note= "dehydratase domain (DH) 6"
FT   misc_feature    complement (16587..17820)
FT                   /*tag= r
FT                   /note= "acyltransferase domain (AT) 6"
```

```
FT   misc_feature    complement (17820..19053)
FT                   /*tag= s
FT                   /note= "KS6"
FT   misc_feature    complement (19116..19326)
FT                   /*tag= t
FT                   /note= "ACP5"
FT   misc_feature    complement (19464..20097)
FT                   /*tag= u
FT                   /note= "KR5"
FT   misc_feature    complement (20241..21420)
FT                   /*tag= v
FT                   /note= "DH5"
FT   misc_feature    complement (21420..22653)
FT                   /*tag= w
FT                   /note= "AT5"
FT   misc_feature    complement (22653..23892)
FT                   /*tag= x
FT                   /note= "KS5"
FT   CDS             complement (23992..46573)
FT                   /*tag= y
FT                   /note= "fkbB gene"
FT   misc_feature    complement (24163..24373)
FT                   /*tag= z
FT                   /note= "ACP4"
FT   misc_feature    complement (24997..26146)
FT                   /*tag= aa
FT                   /note= "DH4 (inactive)"
FT   misc_feature    complement (26146..27430)
FT                   /*tag= ab
FT                   /note= "AT4"
FT   misc_feature    complement (27430..28684)
FT                   /*tag= ac
FT                   /note= "KS4"
FT   misc_feature    complement (28750..28960)
FT                   /*tag= ad
FT                   /note= "ACP3"
FT   misc_feature    complement (29092..29740)
FT                   /*tag= ae
FT                   /note= "KR3"
FT   misc_feature    complement (29869..31018)
FT                   /*tag= af
FT                   /note= "DH3 (inactive)"
FT   misc_feature    complement (31018..32185)
FT                   /*tag= ag
FT                   /note= "AT3"
FT   misc_feature    complement (32185..33439)
FT                   /*tag= ah
FT                   /note= "KS3"
FT   misc_feature    complement (33505..33715)
FT                   /*tag= ai
FT                   /note= "ACP2"
FT   misc_feature    complement (33823..34480)
FT                   /*tag= aj
FT                   /note= "KR2"
FT   misc_feature    complement (34606..35749)
FT                   /*tag= ak
FT   misc_feature    complement (28750..28960)
FT                   /*tag= al
FT                   /note= "DH2 (inactive)"
FT   misc_feature    complement (35749..37144)
FT                   /*tag= am
FT                   /note= "AT2"
FT   misc_feature    complement (37145..38296)
FT                   /*tag= an
FT                   /note= "KS2"
FT   misc_feature    complement (38371..38581)
FT                   /*tag= ao
FT                   /note= "ACP1"
FT   misc_feature    complement (38677..39307)
FT                   /*tag= ap
FT                   /note= "KR1"
FT   misc_feature    complement (39442..40609)
FT                   /*tag= aq
FT                   /note= "DH1"
FT   misc_feature    complement (40609..41842)
FT                   /*tag= ar
FT                   /note= "AT1"
FT   misc_feature    complement (41842..43093)
FT                   /*tag= as
FT                   /note= "KS of extender module 1 (KS1)"
FT   misc_feature    complement (43144..43660)
FT                   /*tag= at
FT                   /note= "ACP of loading domain"
FT   misc_feature    complement (43777..44629)
FT                   /*tag= au
FT                   /note= "ER of loading domain"
FT   misc_feature    complement (44974..46573)
FT                   /*tag= av
FT                   /note= "CoA ligase of laoding domain"
FT   CDS             46754..47788
FT                   /*tag= aw
FT                   /note= "fkbO gene"
FT   CDS             47785..52272
FT                   /*tag= ax
FT                   /note= "fkbP gene"
FT   CDS             52275..71465
FT                   /*tag= ay
FT                   /note= "fkbA gene"
FT   misc_feature    52362..53576
FT                   /*tag= az
FT                   /note= "KS7"
FT   misc_feature    53577..54716
FT                   /*tag= bb
FT                   /note= "AT7"
FT   misc_feature    54717..55871
FT                   /*tag= ba
FT                   /note= "DH7"
FT   misc_feature    56019..56819
FT                   /*tag= bc
FT                   /note= "ER7"
FT   misc_feature    56943..57575
FT                   /*tag= bd
FT                   /note= "KR7"
FT   misc_feature    57710..57920
FT                   /*tag= be
FT                   /note= "ACP7"
FT   misc_feature    57990..59243
FT                   /*tag= bf
FT                   /note= "KS8"
FT   misc_feature    59244..60398
FT                   /*tag= bg
FT                   /note= "AT8"
FT   misc_feature    60399..61412
FT                   /*tag= bh
FT                   /note= "DH8 (inactive)"

  Query Match             52.5%;  Score 47.8;  DB 21;  Length 77536;
  Best Local Similarity   70.3%;  Pred. No. 2.2e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||| | ||  | ||| ||||| |   |||| |  ||||||| || ||| ||||||
Db   53053 CCGACGCCGAGCGCAACGGCCACCGCGTGCTGGCGGTCGTCCGGGGTTCCGCCGTCAACC 53112

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || ||    || | |||| || |||||||
Db   53113 AGGACGGTGCCTCCAACGGGCTGTCCGCGCC 53143


RESULT 12
AAS08693
ID   AAS08693 standard; DNA; 109519 BP.
XX
AC   AAS08693;
XX
DT   26-SEP-2001  (first entry)
```

```
XX
DE   Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.
XX
KW   Everninomicin; antibiotic; bottle-neck gene; orthomicin;
KW   fermentation; ds.
XX
OS   Micromonospora carbonacea var. africana.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (132..1362)
FT                   /*tag= a
FT                   /product= "EvdA"
FT   RBS             complement (1389..1394)
FT                   /*tag= b
FT   CDS             complement (1490..2611)
FT                   /*tag= c
FT                   /product= "EvdB"
FT   RBS             complement (2618..2622)
FT                   /*tag= d
FT   CDS             complement (2622..3860)
FT                   /*tag= e
FT                   /product= "EvdC"
FT   RBS             complement (3867..3870)
FT                   /*tag= f
FT   CDS             4143..5312
FT                   /*tag= g
FT                   /product= "EvdD"
FT   RBS             4134..4138
FT                   /*tag= h
FT   CDS             5309..6235
FT                   /*tag= i
FT                   /product= "EvdE"
FT   CDS             6232..7275
FT                   /*tag= j
FT                   /product= "EvdF"
FT   RBS             6226..6229
FT                   /*tag= k
FT   CDS             7272..8327
FT                   /*tag= l
FT                   /product= "EvdG"
FT   CDS             8342..9364
FT                   /*tag= m
FT                   /product= "EvdH"
FT   RBS             8333..8336
FT                   /*tag= n
FT   CDS             complement (9463..10224)
FT                   /*tag= o
FT                   /product= "EvdI"
FT   RBS             complement (10232..10235)
FT                   /*tag= p
FT   CDS             10424..11176
FT                   /*tag= q
FT                   /product= "EvdJ"
FT   CDS             12027..12455
FT                   /*tag= r
FT                   /product= "EvdK"
FT                   /partial
FT                   /note= "No start codon"
FT   CDS             complement (12108..13022)
FT                   /*tag= s
FT                   /product= "EvdL"
FT   RBS             complement (13027..13030)
FT                   /*tag= t
FT   CDS             complement (14410..15363)
FT                   /*tag= u
FT                   /product= "EvrA"
FT   RBS             complement (15369..15373)
FT                   /*tag= v
FT   CDS             complement (15380..16414)
FT                   /*tag= w
FT                   /product= "EvrB"
FT   CDS             complement 16419..17873
FT                   /*tag= x
FT                   /product= "EvrC"
FT   CDS             complement (17870..18934)
FT                   /*tag= y
FT                   /product= "EvrD"
FT   CDS             19374..20906
FT                   /*tag= z
FT                   /product= "EvrE"
FT   CDS             21064..22542
FT                   /*tag= aa
FT                   /product= "EvrF"
FT   RBS             21056..22542
FT                   /*tag= ab
FT   CDS             22748..24172
FT                   /*tag= ac
FT                   /product= "EvrG"
FT   RBS             22736..22740
FT                   /*tag= ad
FT   CDS             complement (24177..25223)
FT                   /*tag= ae
FT                   /product= "EvrH"
FT   RBS             complement (25230..25233)
FT                   /*tag= af
FT   CDS             25550..26626
FT                   /*tag= ag
FT                   /product= "EvrI"
FT   CDS             26685..30479
FT                   /*tag= ah
FT                   /product= "EvrJ"
FT   RBS             26672..26676
FT                   /*tag= ai
FT   CDS             complement (30557..31876)
FT                   /*tag= aj
FT                   /product= "EvrK"
FT   RBS             complement (31885..31888)
FT                   /*tag= ak
FT   CDS             complement (31941..32882)
FT                   /*tag= al
FT                   /product= "EvrL"
FT   CDS             complement (33167..34405)
FT                   /*tag= am
FT                   /product= "EvrM"
FT   RBS             complement (34414..34418
FT                   /*tag= an
FT   CDS             complement (34449..35210)
FT                   /*tag= ao
FT                   /product= "EvrN"
FT   RBS             complement (35219..35221)
FT                   /*tag= ap
FT   CDS             complement (35294..36238)
FT                   /*tag= aq
FT                   /product= "EvrO"
FT   CDS             complement (36235..36963)
FT                   /*tag= ar
FT                   /product= "EvrP"
FT   CDS             complement (36998..38026)
FT                   /*tag= as
FT                   /product= "EvrQ"
FT   CDS             complement (38072..38566)
FT                   /*tag= at
FT                   /product= "EvrR"
FT   CDS             complement (38892..40163)
FT                   /*tag= au
FT                   /product= "EvrS"
FT   CDS             complement (40216..40890)
FT                   /*tag= av
FT                   /product= "EvrT"
FT   RBS             complement (40899..40902)
FT                   /*tag= aw
FT   CDS             complement (40887..41576)
FT                   /*tag= ax
FT                   /product= "EvrU"
FT   CDS             complement (41679..42707)
FT                   /*tag= ay
```

```
FT                   /product= "EvrV"
FT   RBS             complement (42714..42717)
FT                   /*tag= az
FT   CDS             complement (42810..43799)
FT                   /*tag= ba
FT                   /product= "EvrW"
FT   RBS             complement (43807..43811)
FT                   /*tag= bb
FT   CDS             complement (43799..44866)
FT                   /*tag= bc
FT                   /product= "EvrX"
FT   CDS             complement (45014..45760)
FT                   /*tag= bd
FT                   /product= "EvrY"
FT   RBS             complement (45767..45770)
FT                   /*tag= be
FT   CDS             complement (45962..46714)
FT                   /*tag= bf
FT                   /product= "EvrZ"
FT   RBS             complement (45952..45956)
FT                   /*tag= bg
FT   CDS             complement (47156..49234)
FT                   /*tag= bh
FT                   /product= "EvsA"
FT   CDS             51627..52715
FT                   /*tag= bi
FT                   /product= "EvsB"
FT   RBS             51629..51622
FT                   /*tag= bj
FT   CDS             52889..53557
FT                   /*tag= bk
FT                   /product= "EvsC"
FT   CDS             53554..54207
FT                   /*tag= bl
FT                   /product= "EvbA"
FT   CDS             complement (54362..55117)
FT                   /*tag= bm
FT                   /product= "EvbB"
FT   RBS             complement (55125..55128)
FT                   /*tag= bn
FT   CDS             complement (55135..56094)
FT                   /*tag= bo
FT                   /product= "EvbC"
FT   RBS             complement (56100..56103)
FT                   /*tag= bp
FT   CDS             complement (56184..56813)
FT                   /*tag= bq
FT                   /product= "EvbC2"
FT   CDS             56961..58709

  Query Match             52.5%;  Score 47.8;  DB 22;  Length 109519;
  Best Local Similarity   70.3%;  Pred. No. 2.3e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||| || ||||| || |||||  |   |||  || | ||||||   ||| ||||||
Db  27466 CCGACGCCCAGCGGGACGGGGACCGGGTGCTGGCCGTGCTGCGGGGCAGCGCCGTCAACC 27525

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | || |||||| | | |||| | |  || ||
Db  27526 AGGACGGCCGGACCAACGGGATCATGGCACC 27556


RESULT 13
ABN76829
ID   ABN76829 standard; cDNA; 330 BP.
XX
AC   ABN76829;
XX
DT   08-JUL-2002  (first entry)
XX
DE   Human synthase-like ORF1776 cDNA, SEQ ID NO:3551.
XX
KW   Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW   disease monitoring; cytokine; cell proliferation; cell differentiation;
KW   immune modulation; haematopoiesis regulation; tissue growth;
KW   angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW   behaviour; cancer; proliferative disorder; neurological disorder;
KW   cardiovascular disease; immune system disorder; organ transplantation;
KW   tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW   hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW   vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW   neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW   cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW   dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200190366-A2.
XX
PD   29-NOV-2001.
XX
PF   24-MAY-2001; 2001WO-US17076.
XX
PR   24-MAY-2000; 2000US-206690P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Leach MD,  Shimkets RA;
XX
DR   WPI; 2002-106200/14.
DR   P-PSDB; ABP32803.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and disorders related to organ
PT   transplantation -
XX
PS   Claim 1; Page 1148; 2508pp; English.
XX
CC   Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC   designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC   ABN79587 represent cDNAs encoding them. The invention also encompasses
CC   polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC   referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC   the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC   polynucleotides, the recombinant production of ORFX proteins, antibodies
CC   specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC   polypeptides, methods of screening for modulators of ORFX expression or
CC   activity, and methods of screening individuals for a predisposition to an
CC   ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC   range of biological activities, such as cytokine, cell proliferation,
CC   cell differentiation, immune modulation, haematopoiesis regulation,
CC   tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC   chemokinetic activity, haemostatic activity, thrombolytic activity,
CC   receptor/ligand, antiinflammatory activity, tumour inhibition activity,
CC   and antiinfective activity, and may also be involved in the determination
CC   of bodily characteristics, fertility and behaviour. ORFX proteins,
CC   nucleic acids and antibodies may be used in the treatment of cancers,
CC   other proliferative disorders such as psoriasis and benign tumours,
CC   neurological disorders such as epilepsy and Alzheimer's disease,
CC   cardiovascular diseases, immune system disorders, disorders related to
CC   organ transplantation, disorders of tissue growth and regeneration,
CC   diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC   storage disease, and infectious diseases caused by viral, bacterial,
CC   fungal and other pathogens. ORFX nucleic acids may also be used as a
CC   source of primers and probes, in the detection of ORFX genomic sequences
CC   or transcripts, in the identification and cloning of homologous
CC   sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC   nucleic acids may additionally be used to produce transgenic animals
CC   which may be useful for studying the function and/or activity of ORFX
CC   protein, and in drug screening. The ORFX proteins may also be used as
CC   immunogens to generate specific antibodies, which are useful in the
CC   diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
```

```
SQ   Sequence 330 BP; 57 A; 95 C; 103 G; 70 T; 5 other;

  Query Match            51.0%;  Score 46.4;  DB 24;  Length 330;
  Best Local Similarity  69.7%;  Pred. No. 3.4e-05;
  Matches   62;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy       3 GACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
           ||||||  |  |||| ||||||   |||| |||  | ||    |||||||| |||||
Db     130 GACGCGGAGNGGGACGGCGACGATATCTTGGCTGTCATTAAAGGCTCGGCCGTCAACTCC 189

Qy      63 GATGGCCGGTCGACCGGGTTGACCGCGCC 91
           |||||  |  || |  || |||||||| ||
Db     190 GATGGGCACTCCAATGGCTTGACCGCCCC 218


RESULT 14
AAF89970
ID   AAF89970 standard; DNA; 758 BP.
XX
AC   AAF89970;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Partial nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83962.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 222; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF89964-78
CC   encode partial type I polyketide synthases, and were isolated using the
CC   method of the invention.
XX
SQ   Sequence 758 BP; 121 A; 250 C; 247 G; 140 T; 0 other;

  Query Match            50.8%;  Score 46.2;  DB 22;  Length 758;
  Best Local Similarity  69.2%;  Pred. No. 4.3e-05;
  Matches   63;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| ||||     ||||||||   |||||| ||  | ||||| |||||||| ||||| |
Db     569 CCGATGCGCTGGCCGATGGCGATGCCATCTGTGCAGTCATCCGCGGCTCGGCAATCAATC 628

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || || |||   |  ||  | || |||||
Db     629 AGGACGGACGGAGCAATGGCATCACGGCGCC 659


RESULT 15
AAF90036
ID   AAF90036 standard; DNA; 5292 BP.
XX
AC   AAF90036;
XX
DT   06-AUG-2001  (first entry)
XX
DE   Nucleotide sequence of a type I polyketide synthase.
XX
KW   Metabolic pathway operon; polyketide; polyketide antibiotic;
KW   type I polyketide synthase; ss.
XX
OS   Unidentified.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5292
FT                   /*tag=  a
FT                   /product= "type I polyketide synthase"
XX
PN   WO200140497-A2.
XX
PD   07-JUN-2001.
XX
PF   27-NOV-2000; 2000WO-FR03311.
XX
PR   29-NOV-1999;   99FR-0015032.
PR   07-JUN-2000; 2000US-0209800.
XX
PA   (AVET ) AVENTIS PHARMA SA.
XX
PI   Jeannin P,  Pernodet J,  Guerineau M,  Simonet P,  Courtois S;
PI   Cappellano C,  Francou F,  Raynal A,  Ball M,  Sezonov G,  Tuphile K;
PI   Frostegard A;
XX
DR   WPI; 2001-374849/39.
DR   P-PSDB; AAB83973.
XX
PT   Collection of nucleic acids from environmental samples, useful for
PT   identifying e.g. genes encoding polyketide synthases and derived
PT   antibiotics -
XX
PS   Claim 35; Page 305-307; 356pp; French.
XX
CC   The specification describes a method for the preparation of a collection
CC   of nucleic acids from organisms in a soil sample. The method comprises
CC   milling a dried sample to produce microparticles; suspending these in
CC   liquid buffer; extraction of nucleic acids from the microparticle;
CC   passing nucleic acid-containing solution through a molecular sieve;
CC   passing nucleic acid-enriched fractions through an anion exchange
CC   chromatography material; and recovering fractions containing purified
CC   nucleic acids. The nucleic acids are sources for sequences that encode
CC   either operons involved in a metabolic pathway (specifically polyketide
CC   synthesis) or polypeptides, particularly for production of therapeutic
CC   or agricultural compounds, especially polyketide antibiotics. AAF90034-39
CC   represent open reading frames (ORFS) of the coding strand of cosmid
CC   a26g1, and encode type I polyketide synthases.
XX
SQ   Sequence 5292 BP; 905 A; 1715 C; 1725 G; 947 T; 0 other;

  Query Match            50.8%;  Score 46.2;  DB 22;  Length 5292;
  Best Local Similarity  69.2%;  Pred. No. 5.2e-05;
  Matches   63;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;
```

```
Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||||||||  || ||| |||| |   |   |||| |||| || ||| |||| |||||||
Db     839 CCGACGCGCTGCGCGATCGCGATCCGGTGATGGCGGTGATTCGCGGCACGGCAATCAACC 898

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || || ||    |  ||  |||| |||||
Db     899 AGGACGGACGCAGCAATGGACTGACGGCGCC 929


Search completed: November  5, 2002, 12:32:11
Job time : 232.63 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 5, 2002, 10:31:40 ; Search time 413.361 Seconds
(without alignments)
6406.881 Million cell updates/sec

Title: US-09-724-876-2_COPY_16975_17065
Perfect score: 91
Sequence: 1 ccgacgcgcaacgggatggc.....tcgaccgggttgaccgcgcc 91

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	91	100.0	58733	1	AF217189	AF217189 Sorangium
	2	91	100.0	68750	1	AF210843	AF210843 Sorangium
	3	91	100.0	68750	6	AR193029	AR193029 Sequence
	4	91	100.0	68750	6	AR199551	AR199551 Sequence
	5	91	100.0	68750	6	AR199559	AR199559 Sequence
	6	91	100.0	68750	6	AR199567	AR199567 Sequence
	7	91	100.0	68750	6	AR201097	AR201097 Sequence
	8	91	100.0	68750	6	AR208671	AR208671 Sequence
	9	91	100.0	71989	6	AR172664	AR172664 Sequence
	10	65.4	71.9	16124	1	AX024384	AX024384 Sequence
	11	65.4	71.9	16124	6	AX024277	AX024277 Sequence
c	12	55.8	61.3	49736	1	AF319998	AF319998 Stigmatel
	13	54.2	59.6	104326	1	AB070940	AB070940 Streptomy
	14	52.2	57.4	10910	1	AX024383	AX024383 Sequence
	15	52.2	57.4	10910	6	AX024276	AX024276 Sequence
	16	51	56.0	22838	1	AF081920	AF081920 Pseudomon
	17	51	56.0	27541	6	AX211706	AX211706 Sequence
	18	51	56.0	123580	1	AF263912	AF263912 Streptomy
	19	51	56.0	125401	6	AX211739	AX211739 Sequence
	20	49.4	54.3	30000	6	AX250262	AX250262 Sequence
	21	49.4	54.3	66808	1	SAU421825	AJ421825 Stigmatel
c	22	48.4	53.2	27522	1	AB070942	AB070942 Streptomy
	23	47.8	52.5	671	6	AX153711	AX153711 Sequence
	24	47.8	52.5	764	6	AX153716	AX153716 Sequence
	25	47.8	52.5	3792	6	AX195972	AX195972 Sequence
	26	47.8	52.5	4306	6	AX153796	AX153796 Sequence
	27	47.8	52.5	6462	6	AX153794	AX153794 Sequence
	28	47.8	52.5	34071	6	AX153790	AX153790 Sequence
	29	47.8	52.5	39314	1	SGR300302	AJ300302 Streptomy
c	30	47.8	52.5	42717	6	AX153789	AX153789 Sequence
	31	47.8	52.5	77534	1	AF235504	AF235504 Streptomy
c	32	47.8	52.5	84985	1	SNA278573	AJ278573 Streptomy
	33	47.8	52.5	109519	6	AX195929	AX195929 Sequence
	34	47.8	52.5	113193	1	AF357202	AF357202 Streptomy
	35	46.4	51.0	330	6	AX310566	AX310566 Sequence
	36	46.2	50.8	758	6	AX153712	AX153712 Sequence
	37	46.2	50.8	5292	6	AX153793	AX153793 Sequence
c	38	46.2	50.8	24045	1	AB070941	AB070941 Streptomy
	39	46.2	50.8	30000	6	AX250261	AX250261 Sequence
c	40	46.2	50.8	30000	6	AX250263	AX250263 Sequence
	41	46.2	50.8	32870	1	AE007122	AE007122 Mycobacte
	42	46.2	50.8	42603	1	AF188287	AF188287 Stigmatel
	43	46.2	50.8	65140	6	AX211705	AX211705 Sequence
	44	46.2	50.8	107379	1	SHGCPIR	X86780 S.hygroscop
c	45	46.2	50.8	107379	1	SHGCPIR	X86780 S.hygroscop

ALIGNMENTS

RESULT 1
AF217189
LOCUS AF217189 58733 bp DNA linear BCT 09-JUN-2000
DEFINITION Sorangium cellulosum putative transposase gene, partial cds; putative transposase gene, complete cds; epothilone biosynthesis gene cluster, complete sequence; putative membrane protein gene, complete cds.
ACCESSION AF217189
VERSION AF217189.1 GI:7453554
KEYWORDS .
SOURCE Polyangium cellulosum.
ORGANISM Polyangium cellulosum
Bacteria; Proteobacteria; delta subdivision; Myxobacteria; Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE 1 (bases 1 to 58733)

```
  AUTHORS   Tang,L., Shah,S., Chung,L., Carney,J., Katz,L., Khosla,C. and
            Julien,B.
  TITLE     Cloning and heterologous expression of the epothilone gene cluster
  JOURNAL   Science 287 (5453), 640-642 (2000)
  MEDLINE   20115953
   PUBMED   10649995
REFERENCE   2  (bases 1 to 58733)
  AUTHORS   Julien,B., Shah,S., Ziermann,R., Goldman,R., Katz,L. and Khosla,C.
  TITLE     Isolation and characterization of the epothilone biosynthetic gene
            cluster from Sorangium cellulosum
  JOURNAL   Gene 249 (1-2), 153-160 (2000)
  MEDLINE   20293058
   PUBMED   10831849
REFERENCE   3  (bases 1 to 58733)
  AUTHORS   Julien,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-DEC-1999) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .58733
                     /organism="Polyangium cellulosum"
                     /strain="SMP44"
                     /db_xref="taxon:56'
     CDS             <1. .992
                     /note="orfA"
                     /codon_start=3
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62878.1"
                     /db_xref="GI:7453555"
                     /translation="VRGHVEAFADFGGVPRVLLYDNLKNAVVERHGDAIRFHPTLLAL
                     SADYRFEPRPVAVARGNEKGRVERAIRYVREGFFEARAYADLGDLNRQATEWTSSAAL
                     DRSWVEDRARTVRQAFDDERSVLLRHPDTPFPDHERVEVEVGKTPYARFDLNDYSVPH
                     DRTRRTLVVLADLSQVRIADGNQIVATHVRSWDRGQQIEQPEHLQRLVDEKRRARERR
                     GLDRLARAARSSQAFLRIVAERGDNVGSAIARLLQLLDAVGAAELEEALVEVLERDTI
                     HIGAVRQVIDRRRSERHLPPPVSIPVTRGEHAALVVTPHSLTTYDALKKDPTP"
     CDS             989. .1501
                     /note="orfB"
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="AAF62879.1"
                     /db_xref="GI:7453556"
                     /translation="MTDLTPTETKDRLKSLGLFGLLACWEQLADKPWLREVLAIEERE
                     RHKRSLERRLKNSRVAAFKPMTDFDSSWPKKIDREAVDDLYDSRYADLLFEVVTRRYD
                     AQKPLLLSTNKAFADWGQVFPHAACVVTLVDRLVHRAEVIEIEAESYRLKEAKELNAT
                     RTKQRRTKKH"
     gene            1998. .6263
                     /gene="epoA"
     CDS             1998. .6263
                     /gene="epoA"
                     /note="type I polyketide synthase; epothilone PKS loading
                     module"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoA"
                     /protein_id="AAF62880.1"
                     /db_xref="GI:7453557"
                     /translation="MADPPIERAAEDPIAIVGAGCRLPGGVIDLSGFWTLLEGSRDTV
                     GQVPAERWDAAAWFDPDLDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPRATASAEIDAHGGLGTM
                     PSVGAGRISYVLGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALATDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTAAALDAHAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SSEGLRAALDAAAQGQTPPGVVRGIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWPAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIAAPEA
                     DVAAAVAPHAASVSIAAVNGPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKAGTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACLPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEKG
                     GAVRGGDRRSARLDHPPPESGRREKVEAAGDRPFRLEIDEPGVLDRLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSATEAAAMPVAYLTAWYALDGIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALFEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLT
                     QNLLDALATALSLERVAAENLRAGVQSDFVSSGADQDWEIIAL"
gene                 6260. .10492
                     /gene="epoB"
CDS                  6260. .10492
                     /gene="epoB"
                     /note="non ribosomal peptide synthetase; epothilone PKS
                     module 1"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoB"
                     /protein_id="AAF62881.1"
                     /db_xref="GI:7453558"
                     /translation="MTINQLLNELEHQGVKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVARLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRVAELPPPPMLPMKAD
                     PSTLREIRFRHTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVAFLRRLTEEPWSEQMRCSLPPAQLE
                     ARASANETNSLLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSDAGVEGDGDQLPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAALIEREKVTVWNSVPALMRMLVEHSEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDPSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRNSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERVDARAHAAKADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGLVPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGAALPKFRYPSAG
                     STYPVQTYAYAKSGRIEGVDEGFYYYHPFEHRLLKVSDHGIERGAHVPQNFDVFDEAA
                     FGLLFVGRIDAIESLYGSLSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFDFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGRDQHFADILRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSKDLEQRPNMQDRVEARRKGRRRS"
gene                 10639. .16137
                     /gene="epoC"
CDS                  10639. .16137
                     /gene="epoC"
                     /note="type I polyketide synthase; epothilone PKS module
                     2"
                     /codon_start=1
                     /transl_table=11
                     /product="EpoC"
                     /protein_id="AAF62882.1"
                     /db_xref="GI:7453559"
                     /translation="MEEQESSAIAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQ
                     ELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWE
                     ALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLAT
                     HVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVY
                     AEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNND
                     GARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVF
                     DRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDF
                     ASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAPAAKLPAAAPARSAEL
                     FVVSAKSAAALDAAAARLRDHLQAHQGLSLGDVAFSLATTRSPMEHRLAMAAPSREAL
                     REGLDAAARGQTPPGAVRGRCSPGNVPKVVFVFPGQGSQWVGMGRQLLAEEPVFHAAL
                     SACDRAIQAEAGWSLLAELAADEGSSQIERIDVVQPVLFALAVAFAALWRSWGVGPDV
                     VIGHSMGEVAAAHVAGALSLEDAVAIICRRSRLLRRISGQGEMAVTELSLAEAEAALR
                     GYEDRVSVAVSNSPRSTVLSGEPAAIGEVLSSLNAKGVFCRRVKVDVASHSPQVDPLR
                     EDLLAALGGLRPRAAAVPMRSTVTGAMVAGPELGANYWMNNLRQPVRFAEVVQAQLQG
                     GHGLFVEMSPHPILTTSVEEMRRAAQRAGAAVGSLRRGQDERPAMLEALGALWAQGYP
                     VPWGRLFPAGGRRVPLPTYPWQRERYWIEAPAKSAAGDRRGVRAGGHPLLGEMQTLST
```

```
                     QTSTRLWETTLDLKRLPWLGDHRVQGAVVFPGAAYLEMAISSGAEALGDGPLQITDVV
                     LAEALAFAGDAAVLVQVVTTEQPSGRLQFQIASRAPGAGHASFRVHARGALLRVERTE
                     VPAGLTLSAVRARLQASMPAAATYAELTEMGLQYGPAFQGIAELWRGEGEALGRVRLP
                     DAAGSAAEYRLHPALLDACFQVVGSLFAGGGEATPWVPVEVGSLRLLQRPSGELWCHA
                     RVVNHGRQTPDRQGADFWVVDSSGAVVAEVSGLVAQRLPGGVRRREEDDWFLELEWEP
                     AAVGTAKVNAGRWLLLGGGGGLGAALRSMLEAGGHAVVHAAESNTSAAGVRALLAKAF
                     DGQAPTAVVHLGSLDGGGELDPGLGAQGALDAPRSADVSPDALDPALVRGCDSVLWTV
                     QALAGMGFRDAPRLWLLTRGAQAVGAGDVSVTQAPLLGLGRVIAMEHADLRCARVDLD
                     PTRPDGELGALLAELLADDAEAEVALRGGERCVARIVRRQPETRPRGRIESCVPTDVT
                     IRADSTYLVTGGLGGLGLSVAGWLAERGAGHLVLVGRSGAASVEQRAAVAALEARGAR
                     VTVAKADVADRAQLERILREVTTSGMPLRGVVHAAGILDDGLLMQQTPARFRKVMAPK
                     VQGALHLHALTREAPLSFFVLYASGVGLLGSPGQGNYAAANTFLDALAHHRRAQGLPA
                     LSVDWGLFAEVGMAAAQEDRGARLVSRGMRSLTPDEGLSALARLLESGRVQVGVMPVN
                     PRLWVELYPAAASSRMLSRLVTAHRASAGGPAGDGDLLRRLAAAEPSARSGLLEPLLR
                     AQISQVLRLPEGKIEVDAPLTSLGMNSLMGLELRNRIEAMLGITVPATLLWTYPTVAA
                     LSGHLAREACEAAPVESPHTTADSAVEIEEMSQDDLTQLIAAKFKALT"
     gene            16134. .37907
                     /gene="epoD"
     CDS             16134. .37907
                     /gene="epoD"
                     /note="type I polyketide synthase; epothilone PKS modules
                     3, 4, 5, and 6"
                     /codon_start=1
                     /transl_table=11
                     /product="epoD"
                     /protein_id="AAF62883.1"
                     /db_xref="GI:7453560"
                     /translation="MTTRGPTAQQNPLKQAAIIIQRLEERLAGLAQAELERTEPIAIV
                     GIGCRFPGGADAPEAFWELLDAERDAVQPLDRRWALVGVAPVEAVPHWAGLLTEPIDC
                     FDAAFFGISPREARSLDPQHRLLLEVAWEGLEDAGIPPRSIDGSRTGVFVGAFTADYA
                     RTVARLPREERDAYSATGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRS
                     LRAGESDLALAGGVSTLLSPDMMEAAARTQALSPDGRCRTFDASANGFVRGEGCGLVV
                     LKRLSDAQRDGDRIWALIRGSAINHDGRSTGLTAPNVLAQETVLREALRSAHVEAGAV
                     DYVETHGTGTSLGDPIEVEALRATVGPARSDGTRCVLGAVKTNIGHLEAAAGVAGLIK
                     AALSLTHERIPRNLNFRTLNPRIRLEGSALALATEPVPWPRTDRPRFAGVSSFGMSGT
                     NAHVVLEEAPAVELWPAAPERSAELLVLSGKSEGALDAQAARLREHLDMHPELGLGDV
                     AFSLATTRSAMTHRLAVAVTSREGLLAALSAVAQGQTPAGAARCIASSSRGKLALLFT
                     GQGAQTPGMGRGLCAAWPAFREAFDRCVTLFDRELDRPLREVMWAEAGSAESLLLDQT
                     AFTQPALFAVEYALTALWRSWGVEPELLVGHSIGELVAACVAGVFSLEDGVRLVAARG
                     RLMQGLSAGGAMVSLGAPEAEVAAAVAPHAAWVSIAAVNGPEQVVIAGVEQAVQAIAA
                     GFAARGVRTKRLHVSHAFHSPLMEPMLEEFGRVAASVTYRRPSVSLVSNLSGKVVTDE

  Query Match          100.0%;  Score 91;  DB 1;  Length 58733;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   16975 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 17034

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           |||||||||||||||||||||||||||||||
Db   17035 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 17065


RESULT 2
AF210843
LOCUS       AF210843               68750 bp    DNA     linear   BCT 21-JAN-2000
DEFINITION  Sorangium cellulosum strain So ce90 epothilone biosynthesis gene
            cluster, complete sequence.
ACCESSION   AF210843
VERSION     AF210843.1  GI:6724237
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Molnar,I., Schupp,T., Ono,M., Zirkle,R., Milnamow,M.,
            Nowak-Thompson,B., Engel,N., Toupet,C., Stratmann,A., Cyr,D.D.,
            Gorlach,J., Mayo,J.M., Hu,A., Goff,S., Schmid,J. and Ligon,J.M.
  TITLE     The biosynthetic gene cluster for the microtubule-stabilizing
            agents epothilones A and B from Sorangium cellulosum So ce90
  JOURNAL   Chem. Biol. 7 (2), 97-109 (2000)
  MEDLINE   20130945
   PUBMED   10662695
REFERENCE   2  (bases 1 to 68750)
  AUTHORS   Molnar,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Natural Product Genetics, Novartis
            Agribusiness Research Institute, Inc., 3054 Cornwallis Rd, P.O.Box
            12257, Research Triangle Park, NC 27709, USA
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="Polyangium cellulosum"
                     /strain="So ce90"
                     /db_xref="taxon:56"
     CDS             <1. .1826
                     /note="ORF1"
                     /codon_start=3
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26904.1"
                     /db_xref="GI:6724238"
                     /translation="ASLDALFARATSARVLDDGHGRATERHVLAEARGIEDLRALREH
                     LRIQEGGPSFHCMCLGDLTVELLAHDQPLASISFHHARSLRHPDWTSDAMLVDGPALV
                     RWLAARGAPGPLREYEEERERARTAQEARRLWLAAAPPCFAPDLPRFEDDANGLPLGP
                     MSPEVAEAERRLRASYATPELACAALLAWLGTGAGPWSGYPAYEMLPENLLLGFGLPT
                     AIAAASAPGTSEAALRGAARLFASWEVVSSKKSQLGNIPEALWERLRTIVRAMGNADN
                     LSRFERAEAIAAEVRRLRAQPAPFAAGAGLAVAGVSSSGRLSGLVTDGDALYSGDGND
                     IVMFQPGRISPVVLLAGTDPFFELAPPLSQMLFVAHANAGTISKVLTEGSPLIVMARN
                     QARPMSLVHARGFMAWVNQAMVPDPERGAPFVVQRSTIMEFEHPTPRCLHEPAGSAFS
                     LACDEEHLYWCELSAGRLELWRHPHHRPGAPSRFAYLGEHPIAATWYPSLTLNATHVL
                     WADPDRRAILGVDKRTGVEPIVLAETRHPPAHVVSEDRDIFALTGQPDSRDWHVEHIR
                     SGASTVVADYQRQLWDRPDMVLNRRGLFFTTNDRILTLARS"
     CDS             complement(1900. .3171)
                     /note="ORF2; similar to hypothetical proteins from
                     Mycobacterium tuberculosis and Streptomyces coelicolor and
                     to DD-peptidases"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26905.1"
                     /db_xref="GI:6724239"
                     /translation="MGALISVAAPGCALGGAEEEGQPGQDAGAGALAPAREVMAAEVA
                     AGQMPGAVWLVARGDDVHVDAVGVTELGGSAPMRRDTIFRIASMTKAVTATAVMMLVE
                     EGKLDLDSPVDRWLPELANRKVLARIDGPIDETVPAERPITVRDLMTFTMGFGISFDA
                     SSPIQRAIDELGLVNAQPVPMTPHGPDEWIRRLGTLPLMHQPGAQWMYNTGSLVQGVL
                     VGRAADQGFDAFVRERILAPLGMRDTDFHVPADKLARFAGCGYFTDEQTGEKTRMDRD
                     GAESAYASPPAFPSGAAGLVSTVDDYLLFARMLMNGGVHEGRRLLSAASVREMTADHL
                     TPAQKAASSFFPGFFETHGWGYGMAVVTAPDAVSEVPGRYGWDGGFGTSWINDPGREL
                     IGIVMTQSAGFLFSGALERFWRSVYVATESA"
     CDS             3415. .5556
                     /note="ORF3; similar to Na/H and drug/H antiporters"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26906.1"
                     /db_xref="GI:6724240"
                     /translation="MHGLTERQVLLSLVTLALILVTARASGELARRLRQPEVLGELFG
                     GVVLGPSVVGALAPGFHRALFQEPAVGVVLSGISWIGALLLLLMAGIEVDVGILRKEA
                     RPGALSALGAIAPPLAAGAAFSALVLDRPLPSGLFLGIVLSVTAVSVIAKVLIERESM
                     RRSYAQVTLAAGVVSEVAAWVLVAMTSSSYGASPALAVARSALLASGFLLFMVLVGRR
                     LTHLAMRWVADATRVSKGQVSLVLVLTFLAAALTQRLGLHPLLGAFALGVLLNSAPRT
                     NRPLLDGVQTLVAGLFAPVFFVLAGMRVDVSQLRTPAAWGTVALLLATATAAKVVPAA
                     LGARLGGLRGSEAALVAVGLNMKGGTDLIVAIVGVELGLLSNEAYTMYAVVALVTVTA
                     SPALLIWLEKRAPPTQEESARLEREEAARRAYIPGVERILVPIVAHALPGFATDIVES
                     IVASKRKLGETVDITELSVEQQAPGPSRAAGEASRGLARLGARLRVGIWRQRRELRGS
                     IQAILRASRDHDLLVIGARSPARARGMSFGRLQDAIVQRAESNVLVVVGDPPAAERAS
                     ARRILVPIIGLEYSFAAADLAAHVALAWDAELVLLSSAQTDPGAVVWRDREPSRVRAV
                     ARSVVDEAVFRGRRLGVRVSSRVHVGAHPSDEITRELARAPYDLLVLGCYDHGPLGRL
                     YLGSTVESVVVRSRVPVALLVAHGGTREQVR"
     CDS             complement(5612. .5992)
                     /note="ORF4"
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="unknown"
                     /protein_id="AAF26907.1"
                     /db_xref="GI:6724241"
                     /translation="MDKPIGRTRCAIAEGYIPGGSNGPEPQMTSHETACLLNASDRDA
                     QVAITVYFSDRDPAGPYRVIVPARRTRHVRFNDLTEPEPIPRDTDYASVIESDAPIVV
                     QHTRLDSRQAENALLSTIAYTDRE"
CDS                  6226. .6675
                     /note="ORF5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAF26908.1"
                     /db_xref="GI:6724242"
                     /translation="MKHVETGRRFGRRIGHTLGLLASMALAGCGGPSEKTVQGTRLAP
                     GADARVTADVDPDAATTRLAVDVVHLSPPERLEAGSERFVVWQRPSPESPWRRVGVLD
                     YNADSRRGKLAETTVPYANFELLITAEKQSSPQSPSSAAVIGPTSVG"
gene                 7610. .11875
                     /gene="epoA"
CDS                  7610. .11875
                     /gene="epoA"
                     /note="EPOS A"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26919.1"
                     /db_xref="GI:6724253"
                     /translation="MADRIERAAEDPIAIVGASCRLPGGVIDLSGFWTLLEGSRDTV
                     GRVPAERWDAAAWFDPDPDAPGKTPVTRASFLSDVACFDASFFGISPREALRMDPAHR
                     LLLEVCWEALENAAIAPSALVGTETGVFIGIGPSEYEAALPQATASAEIDAHGGLGTM
                     PSVGAGRISYALGLRGPCVAVDTAYSSSLVAVHLACQSLRSGECSTALAGGVSLMLSP
                     STLVWLSKTRALARDGRCKAFSAEADGFGRGEGCAVVVLKRLSGARADGDRILAVIRG
                     SAINHDGASSGLTVPNGSSQEIVLKRALADAGCAASSVGYVEAHGTGTTLGDPIEIQA
                     LNAVYGLGRDVATPLLIGSVKTNLGHPEYASGITGLLKVVLSLQHGQIPAHLHAQALN
                     PRISWGDLRLTVTRARTPWPDWNTPRRAGVSSFGMSGTNAHVVLEEAPAATCTPPAPE
                     RPAELLVLSARTASALDAQAARLRDHLETYPSQCLGDVAFSLATTRSAMEHRLAVAAT
                     SREGLRAALDAAAQGQTSPGAVRSIADSSRGKLAFLFTGQGAQTLGMGRGLYDVWSAF
                     REAFDLCVRLFNQELDRPLREVMWAEPASVDAALLDQTAFTQPALFTFEYALAALWRS
                     WGVEPELVAGHSIGELVAACVAGVFSLEDAVFLVAARGRLMQALPAGGAMVSIEAPEA
                     DVAAAVAPHAASVSIAAVNAPDQVVIAGAGQPVHAIAAAMAARGARTKALHVSHAFHS
                     PLMAPMLEAFGRVAESVSYRRPSIVLVSNLSGKACTDEVSSPGYWVRHAREVVRFADG
                     VKALHAAGAGTFVEVGPKSTLLGLVPACMPDARPALLASSRAGRDEPATVLEALGGLW
                     AVGGLVSWAGLFPSGGRRVPLPTYPWQRERYWIDTKADDAARGDRRAPGAGHDEVEEG
                     GAVRGGDRRSARLDHPPPPESGRREKVEAAGDRPFRLEIDEPGVLDHLVLRVTERRAPG
                     LGEVEIAVDAAGLSFNDVQLALGMVPDDLPGKPNPPLLLGGECAGRIVAVGEGVNGLV
                     VGQPVIALSAGAFATHVTTSAALVLPRPQALSAIEAAAMPVAYLTAWYALDRIARLQP
                     GERVLIHAATGGVGLAAVQWAQHVGAEVHATAGTPEKRAYLESLGVRYVSDSRSDRFV
                     ADVRAWTGGEGVDVVLNSLSGELIDKSFNLLRSHGRFVELGKRDCYADNQLGLRPFLR
                     NLSFSLVDLRGMMLERPARVRALLEELLGLIAAGVFTPPPIATLPIARVADAFRSMAQ
                     AQHLGKLVLTLGDPEVQIRIPTHAGAGPSTGDRDLLDRLASAAPAARAAALEAFLRTQ
                     VSQVLRTPEIKVGAEALFTRLGMDSLMAVELRNRIEASLKLKLSTTFLSTSPNIALLA
                     QNLLDALATALSLERVAAENLRAGVQNDFVSSGADQDWEIIAL"
misc_feature         7640. .8920
                     /gene="epoA"
                     /note="Region: beta-ketoacyl synthase/malonyl-CoA
                     decarboxylase"
misc_feature         9236. .10201
                     /gene="epoA"
                     /note="Region: acetate acyltransferase"
misc_feature         10529. .11428
                     /gene="epoA"
                     /note="Region: encyl reductase"
misc_feature         11549. .11764
                     /gene="epoA"
                     /note="Region: acyl carrier protein"
gene                 11872. .16104
                     /gene="epoP"
CDS                  11872. .16104
                     /gene="epoP"
                     /note="EPOS P; contains condensation/heterocyclization,
                     cysteine adenylation, putative NAD(P)H oxidase, and
                     peptidyl carrier domains"
                     /codon_start=1
                     /transl_table=11
                     /product="nonribosomal peptide synthetase"
                     /protein_id="AAF26925.1"
                     /db_xref="GI:6724259"
                     /translation="MTINQLLNELEHQGIKLAADGERLQIQAPKNALNPNLLARISEH
                     KSTILTMLRQRLPAESIVPAPAERHAPFPLTDIQESYWLGRTGAFTVPSGIHAYREYD
                     CTDLDVPRLSRAFRKVVARHDMLRAHTLPDMMQVIEPKVDADIEIIDLRGLDRSTREA
                     RLVSLRDAMSHRIYDTERPPLYHVVAVRLDERQTRLVLSIDLINVDLGSLSIIFKDWL
                     SFYEDPETSLPVLELSYRDYVLALESRKKSEAHQRSMDYWKRRIAELPPPPTLPMKAD
                     PSTLKEIRFRHTEQWLPSDSWGRLKRRVGERGLTPTGVILAAFSEVIGRWSASPRFTL
                     NITLFNRLPVHPRVNDITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVS
                     GIEVQREAARVLGIQRGALFPVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDH
                     QLYEHDGDLVLAWDIVDGVFPPDLLDDMLEAYVVFLRRLTEEPWGEQVRCSLPPAQLE
                     ARASANATNALLSEHTLHGLFAARVEQLPMQLAVVSARKTLTYEELSRRSRRLGARLR
                     EQGARPNTLVAVVMEKGWEQVVAVLAVLESGAAYVPIDADLPAERIHYLLDHGEVKLV
                     LTQPWLDGKLSWPPGIQRLLVSEAGVEGDGDQPPMMPIQTPSDLAYVIYTSGSTGLPK
                     GVMIDHRGAVNTILDINERFEIGPGDRVLALSSLSFDLSVYDVFGILAAGGTIVVPDA
                     SKLRDPAHWAELIEREKVTVWNSVPALMRMLVEHFEGRPDSLARSLRLSLLSGDWIPV
                     GLPGELQAIRPGVSVISLGGATEASIWSIGYPVRNVDLSWASIPYGRPLRNQTFHVLD
                     EALEPRPVWVPGQLYIGGVGLALGYWRDEEKTRKSFLVHPETGERLYKTGDLGRYLPD
                     GNIEFMGREDNQIKLRGYRVELGEIEETLKSHPNVRDAVIVPVGNDAANKLLLAYVVP
                     EGTRRRAAEQDASLKTERIDARAHAAEADGLSDGERVQFKLARHGLRRDLDGKPVVDL
                     TGQDPREAGLDVYARRRSVRTFLEAPIPFVEFGRFLSCLSSVEPDGATLPKFRYPSAG
                     STYPVQTYAYVKSGRIEGVDEGFYYYHPFEHRLLKLSDHGIERGAHVRQNFDVFDEAA
                     FNLLFVGRIDAIESLYGSSSREFCLLEAGYMAQLLMEQAPSCNIGVCPVGQFNFEQVR
                     PVLDLRHSDVYVHGMLGGRVDPRQFQVCTLGQDSSPRRATTRGAPPGREQHFADMLRD
                     FLRTKLPEYMVPTVFVELDALPLTSNGKVDRKALRERKDTSSPRHSGHTAPRDALEEI
                     LVAVVREVLGLEVVGLQQSFVDLGATSIHIVRMRSLLQKRLDREIAITELFQYPNLGS
                     LASGLRRDSRDLDQRPNMQDRVEVRRKGRRRS"
gene                 16251. .21749
                     /gene="epoB"
CDS                  16251. .21749
                     /gene="epoB"
                     /note="EPOS B"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAF26920.1"
                     /db_xref="GI:6724254"

  Query Match            100.0%;  Score 91;  DB 1;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           |||||||||||||||||||||||||||||||
Db   22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 3
AR193029
LOCUS       AR193029               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6346404.
ACCESSION   AR193029
VERSION     AR193029.1  GI:20238994
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6346404-A 1 12-FEB-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN
```

```
  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 4
AR199551
LOCUS       AR199551               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355457.
ACCESSION   AR199551
VERSION     AR199551.1  GI:20249625
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355457-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 5
AR199559
LOCUS       AR199559               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355458.
ACCESSION   AR199559
VERSION     AR199559.1  GI:20249633
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355458-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 6
AR199567
LOCUS       AR199567               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6355459.
ACCESSION   AR199567
VERSION     AR199567.1  GI:20249641
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6355459-A 1 12-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 7
AR201097
LOCUS       AR201097               68750 bp    DNA     linear   PAT 20-APR-2002
DEFINITION  Sequence 1 from patent US 6358719.
ACCESSION   AR201097
VERSION     AR201097.1  GI:20251985
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6358719-A 1 19-MAR-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
```

```
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 8
AR208671
LOCUS       AR208671               68750 bp    DNA     linear   PAT 20-JUN-2002
DEFINITION  Sequence 1 from patent US 6383787.
ACCESSION   AR208671
VERSION     AR208671.1  GI:21509886
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 68750)
  AUTHORS   Schupp,T., Ligon,J.Madison., Molnar,I., Zirkle,R., Cyr,D.Dawn. and
            Gorlach,J.
  TITLE     Genes for the biosynthesis of epothilones
  JOURNAL   Patent: US 6383787-A 1 07-MAY-2002;
FEATURES             Location/Qualifiers
     source          1. .68750
                     /organism="unknown"
BASE COUNT     9596 a  22456 c  25539 g  11159 t
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 68750;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22587 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 22646

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  22647 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 22677


RESULT 9
AR172664
LOCUS       AR172664               71989 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 2 from patent US 6303342.
ACCESSION   AR172664
VERSION     AR172664.1  GI:17912155
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 71989)
  AUTHORS   Julien,B., Katz,L., Khosla,C. and Tang,L.
  TITLE     Recombinant methods and materials for producing epothilones C and D
  JOURNAL   Patent: US 6303342-A 2 16-OCT-2001;
FEATURES             Location/Qualifiers
     source          1. .71989
                     /organism="unknown"
BASE COUNT    10108 a  23531 c  26617 g  11731 t      2 others
ORIGIN

  Query Match           100.0%;  Score 91;  DB 6;  Length 71989;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-12;
  Matches   91;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  16975 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 17034

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          |||||||||||||||||||||||||||||||
Db  17035 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 17065


RESULT 10
AX024384
LOCUS       AX024384               16124 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024384
VERSION     AX024384.1  GI:10184588
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match           71.9%;  Score 65.4;  DB 1;  Length 16124;
  Best Local Similarity  82.4%;  Pred. No. 2.5e-06;
  Matches   75;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||| || ||  | ||||| || |||||||| |||||||||| |||||||| || |
Db    472 CCGACGCCCAGCGACACGGCGATCGGATCTGGGCTCTGATCCGGGGTTCGGCCATGAATC 531

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | ||||||||||||| |||||||  || ||
Db    532 AGGATGGCCGGTCGACAGGGTTGATGGCACC 562


RESULT 11
AX024277
LOCUS       AX024277               16124 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 82 from Patent DE19846493.
ACCESSION   AX024277
VERSION     AX024277.1  GI:10184551
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 16124)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 82 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .16124
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     2336 a   5400 c   5695 g   2693 t
ORIGIN

  Query Match           71.9%;  Score 65.4;  DB 6;  Length 16124;
  Best Local Similarity  82.4%;  Pred. No. 2.5e-06;
  Matches   75;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy      1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
          ||||||| || ||  | ||||| || |||||||| |||||||||| |||||||| || |
Db    472 CCGACGCCCAGCGACACGGCGATCGGATCTGGGCTCTGATCCGGGGTTCGGCCATGAATC 531

Qy     61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
          | ||||||||||||| |||||||  || ||
Db    532 AGGATGGCCGGTCGACAGGGTTGATGGCACC 562


RESULT 12
AF319998/c
LOCUS       AF319998               49736 bp    DNA     linear   BCT 28-MAY-2001
DEFINITION  Stigmatella aurantiaca myxalamid biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF319998
```

```
VERSION     AF319998.1  GI:14210834
KEYWORDS    .
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Kunze,B., Blocker,H. and Muller,R.
  TITLE     Novel features in a combined polyketide synthase/non-ribosomal
            peptide synthetase: the myxalamid biosynthetic gene cluster of the
            myxobacterium Stigmatella aurantiaca Sga15
  JOURNAL   Chem. Biol. 8 (1), 59-69 (2001)
  MEDLINE   21110452
   PUBMED   11182319
REFERENCE   2  (bases 1 to 49736)
  AUTHORS   Silakowski,B., Nordsiek,G., Bloecker,H. and Mueller,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2000) MX, GBF, Mascheroderweg 1, Braunschweig
            38124, Germany
FEATURES             Location/Qualifiers
     source          1. .49736
                     /organism="Stigmatella aurantiaca"
                     /db_xref="taxon:41"
     CDS             complement(2. .1306)
                     /note="ORF1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative aspartokinase"
                     /protein_id="AAK57182.1"
                     /db_xref="GI:14210835"
                     /translation="MRPIGSARQAGRGTSALALIVQKYGGTSVGDTERIKNVAKRCIA
                     AQKAGHDVVVVVSAMSGETNRLLKLVSQITERPNEREQDVVVATGEQVSIGLVALAIQ
                     TQKRKAVSFLGHQVRITTDSTFSKARIKSIDAERIVDALKKKNIVVVAGFQGQDEHGN
                     VTTLGRGGSDTTAVALAAALKADACEIYTDVDGVYTTDPNMVPAARKLERIAYEEMLE
                     LASVGAKVLQIRSVEFAMKYKVPLWVKSSFTDDPGTLVCEEDAAMEDVLVSGIAYDKN
                     EAKIAIRGVPDVPGVAAKIFGALDEQSIVVDLIVQNVSKDGRTDVTFTVGKADLTKAK
                     DVVKKIAKSVKAEGVETDDQVSKVSIVGVGMRNHSGVAAKMFTVLAGAGVNVQMISTS
                     EIKVSCVIHSNYTELAVRSLHTAFGLDKPEAGAVSEVVSLKG"
     CDS             1363. .2895
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative histidine ammonium lyase"
                     /protein_id="AAK57183.1"
                     /db_xref="GI:14210836"
                     /translation="MSRPRLNIDGDTLKLEEILQVARHTVTVELAPAAAARVKAARDL
                     VDRVAAGDTPSYGINTGFGTLAEVRIDKKDLRELQRNLILSHAAGVGSPLPLPEARVL
                     LLLRCNVLAKGYSGIRPETLALALEMLNRDVVPVVPERGSVGASGDLAPLAHLALVFI
                     GEGEAFYKGERLPAAQALERAGLKPVVLEAKEGLALVNGTQAMCAVGTLLQLRAEMLA
                     DLADLAGAMTLEGLLGSHKPFIPEIQDVRAHEGQKACAAHLRELLADSALVESHVNCS
                     KVQDPYSLRCMPQVHGAAREGLSFARRILEVEINSATDNPLVFVETERIVSGGNFHGQ
                     PVSLALDVAAMALTQLSAISERRVEQLVNPALSGLPPFLAKNSGLNSGFMIAQVTSAA
                     LVAESRVLSHPASVDSIPSSAGREDHVSMGMTAALKGRQVADFTRSCLAIELLVAAQA
                     LDYRQPTRAGKGPQAAYELIRSKIPTMEKDRELHRDIAAVSALIDSGELLNAVRTATR
                     GQ"
     gene            complement(3275. .7822)
                     /gene="mxaA"
     CDS             complement(3275. .7822)
                     /gene="mxaA"
                     /note="non-ribosomal peptide synthetase; contains a
                     condensing, adenylation, and PCP reductase domains"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaA"
                     /protein_id="AAK57184.1"
                     /db_xref="GI:14210837"
                     /translation="MSVRELLTTLAAKGVRLRLEGDKLQVEASKAVLTPELRAELSAN
                     KADILAFLRKHSSREETGSELGQPLARPEVIPLSSGQERLWFLDRLSPGTSQYNVHLG
                     LRARGALDTKALRRSLDELVRRHEVLRTHFPEVEGSPRQVIVSPDTGVALTIVELPGL
                     TESQREEEIYRRSDELAQQPFDLAKGPLLRVTVLALGPDDFALFVTKHHIITDGWSLG
                     VFVRELSMLYASFVRGQAAALPPVSMQFADSALRERAWLAGESGARERAYWKGKLKGL
                     PPLQLPVDHAVSTSTSHRGATVPLALSPALSDALRELASREGCSFFTVLFAAFSALLH
                     RYSGQVDFGVGTVIANRGSVPADLIGFIANTLALRCDLSGEPTFSQWLARAHKVVLEA
                     LDYQALPFSEVVQTVGASRDGGLNPLVRACFTLESIPAPTLDLPGTSWSFLNGAPDGS
                     VEGVAKFELSLILAASEKGLAGMLEYSREVFDASTVERMVGHFQVLLESIVAHPEVPL
                     SKLPLLTAEERGRLLNDWNGPVLDVPALCMHELVQAQAERTPQAVAVVSGQKTLTYAE
                     LNRRANQLAHHLRRLGIQKEERVGLCVERTEDIVIGLLAILKAGGAYVPLDPAYPKER
                     LALILEDAQVPVLLTQQRLVPELPATQARVVCLDKDWPTIGAESDTNPERITAPEAIA
                     YLIYTSGSTGKPKGVMIEHRNAVAFLIWAMSVFSPKELAGTLASTSICFDLSVFEIFT
                     PLCCGAKVIVAKNALELPELPAAREVTLINTVPSAMGALLRSGGVPSSVAIVNLAGEA
                     LAGALVDQIYQLDHVRDVFNLYGPSETTTYSTFTRVNRGQTPTIGRPVGNTQVYALDS
                     NREPMPIGVPGEVYIGGMGVARGYLGRPELTAERFVRSPFGGGSEARLYRTGDLARWL
                     PDGQLEYLGRMDHQVKLRGFRIELGEIGAVLMEHSGIRDAVVVVREGLGADKQLVAYV
                     VGRGEKAPEPAELRDYLKSKLPEYMVPFLFVGLDALPLTPNGKVDRAALPAPERTHSG
                     PAKEHVAPRTPGEESLAAIWRQVLGVEQIGAHDNFFELGGHSLLLYRVLVLARSASGA
                     DIPLRALLQAPTLEEMARAVEAAKTGSLPAHDVTVEMEADAVLDAEIALGKALPPVTG
                     ALRTILLTGATGFLGAFLLEELCRRTDARIYCLVRSKTEQEGMNRIKNLESYSLWNE
                     ALAPRIVPVRGDIGQPLLGLSEKEFQRLSEEIDAIYHNGALVNFLYPYESMRAANVLG
                     TREILRLATRTRIKPLHYVSTVSVLPLGRKAPIREDEPLEGPSSLVGGYAQSKWVAEK
                     LVREASRRGLPVTILRPGRVTGHSRTGAWNTDDLVCRTLKGCVRMGVAPSVDALLDLT
                     PVDYVSSAIVDLSMRPESIGQTYHLVNPQFVRADEMWNYMRAFGYGLRVLPYDQWLSE
                     LGSAASSDSELGDLLMFLQQVPPEDRSVGGPRMVVCDSGDTLKALGGTGTSCPSVDAS
                     LISTYLSSLVHRGFLKAPEVR"
gene                 complement(7819. .11520)
                     /gene="mxaB1"
CDS                  complement(7819. .11520)
                     /gene="mxaB1"
                     /note="PKS; ketosynthase; acyltransferase"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaB1"
                     /protein_id="AAK57185.1"
                     /db_xref="GI:14210838"
                     /translation="MRAGAGGHPLLGVSFTTSVQKENRFWESALSAKEPSYLADHRVG
                     GAVVVPGAAYLEMALSAARELSGEKAYELVDTSFKEGLLLPEGQERVVQMALSGEGGD
                     ALSFQLSSQKASTEPNAPAGWMAHVVGQIRPIEGGPSTQTPESFEAIRERCSEVIAQA
                     AHYEELASRGVAYGPAFQGVQQIWRGRAEALGHVRLPEPLAARASAYRIHPALLDACF
                     QLVAAAVPQEDLTPDAGPAVPVALAHLRLHEQPGTEVFCHVRLRPSEGARRGVYEFDL
                     VLRDASGRILVEALGLRVQQLDAPAAEHAGKDLFFALEWRHASAVQQVPAEGPGASGR
                     WLLIADGSELADTVESQLRERHGEVIRVDLAGAGQGRRVVDATSPQAFDGVLSEAFGA
                     GSAPRGVIHLTGLEPREVDPIMGCGGVLHLVQALSRVSLASPPRLWLVTRGVHHGSGS
                     RSASDVVQASLWGLGRTLAHEHSELKCTRVDVSAAMGTGDAGAALMRELLVADDEEEI
                     SLRADGRHVGRIVRGSPGRAPSEAVVPAGGRPFHLERDAQGRLEFRAGARRPPGPGEL
                     EIEVEVAALGGPGEGTGPERWEGAGEGIGCSGRVVAIGEGVKNVAVGEERIALVASGL
                     GSHVSVPAGCTAIRPSSSSHGEAAALAATVLPAWYALFHLGRLEKRGRVWVLGATSGL
                     GRAAIKLAKRAGAEVFAVATTDEQRASLRELGADQVMDPSDASFVSRVLDATGGRGVE
                     LAIVAPGAVELERSLPTLSEGARVLDLRASGASARTGDANVAWCVVDVQDFARRRPER
                     FASLWQEVKAAFEAGGLPSAAADAKALALKDPKARVVVPAADARRLRADGTYLVTGGL
                     GGLGLAVAKWMVEQGAGHLVLVGRDTVLSPEQQEVVTALEAAGARVRVARADVSDRAQ
                     IARILSEIAEGGAPLRGVIHAAGVLDDGVLLQQTVERFRRVMAPKVLGGWNLHVLTRE
                     APLDFFVLYSSAASMFGAPGQGNYVAGNAFLDALAHHRRALGLPGLSINWGPFSEVGL
                     AAAQANRGERLAQRGSDSLTPAEGNTILGRLLDGDVTQMAVMPLELRKWVEFYPRAKS
                     SPWLSELVPADAGQEVGEARDVALLETLRTATPQEAKAALEQFVREQLARVLRLDSAR
                     IDPEAPLQSFGLDSLMGLELRNRLVSGLGVSLPASLIWKHPTLAALCAHLLGEVMDRT
                     LAETLAQAAAGAGPAVDESAADDKEVFVL"
gene                 complement(11487. .14303)
                     /gene="mxaB2"
CDS                  complement(11487. .14303)
                     /gene="mxaB2"
                     /note="PKS; dehydratase; enoylreductase; ketoreductase;
                     ACP"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaB2"
                     /protein_id="AAK57186.1"
                     /db_xref="GI:14210839"
                     /translation="MSKQEKPAEQPSALQRAALLVEKMQARLDAVERARTEPIAVIGL
                     GCRFPGGAVDGSSYWKILRDGVDALREVPESRWDVPGHFDSTRGVPGKMYGTRGGFLD
                     DVEHFDAEFFGISPREAASLDPQQRLVLEVAWEALENAGLAPDQLAGSKTGVFMGVMS
                     SDYMARLLKENDATRFDGYMATGNGYSFVPGRVSYVLGLQGPCMPVDTACSSSLVALH
                     LASESLRAGESNLALAGGVNLILSPETMICLCSMQALASDGRCKTFDASADGYVRGEG
                     CGVLVLKRLSDAQRDGDSILALIRGSAVNHDGASGGLTVPNGPSQQAVVQRALDNARI
                     APALVGYIEAHGTGTPLGDPIELRALGAVLGKGRPADRPFFIGSVKTNIGHLEPAAGI
                     AGVIKTILSLQHKEIPPHLHFRTPNPHVEWDRIPARVPVERVPWPAHEGRRIAGVSSF
                     GLSGINAHVVLEEAPASTAPTAVAEEDKAQLVVLSAKTPRALSSLAEAWSSFLKNEEG
                     GSLADISYTSALSRTPHEHRLALAAKSKQELIEQLEAFAAGKSHAGLSSGRRGTGLGQ
```

```
                     KIVFVFPGQGSQWLGMGQRLYAENVTFREAIDWCHEAISRYTDWSLRDLLTNPGQQPR
                     WNDIDVIQPTLFALQVALAALWRWLGLEPHAVVGHSMGEVAAAHVAGALSLEDAARII
                     CERSKLLRRVSGQGAMAVVDLSREQAEVEIRGAEDRIAIAVSNGPRSTVLSGDPAALK
                     EVLERLERREVFCRWVKVDVASHSPQMDPLRQDLLDRMAAIRPTGSAVPLYSTVTGAP
                     IDGRELSAPYWVRNLRQPVLFADAVQRLIQDGNVLFIELSPHPILVPFVDAMLREGDT
                     SGRGLVVPSLIRDQEEWPSLLSSVGALFTRVIAWTGSGYIPTSVAGLRFLRTRGSASG
                     TGWSCQQLRRRAGSCALAREVIPS"
     gene            complement(14300. .29878)
                     /gene="mxaC"
     CDS             complement(14300. .29878)
                     /gene="mxaC"
                     /note="PKS; modules 1,2 and 3 contain KS, AT, DH, KR, and
                     ACP regions"
                     /codon_start=1
                     /transl_table=11
                     /product="MxaC"
                     /protein_id="AAK57187.1"
                     /db_xref="GI:14210840"
                     /translation="MANRNDESNELSPAKRMLVALEKMQTRLNAVEGAAKEPIAIVGM
                     ACRFPQADNVEAFWRLLQGGVDAVREIPLERWPKDEVADIGDGAMRWGGFLDSVDSFD
                     PGFFGISPREAVRMDPQQRLLLEVAWEALEDAGLDVDKLSGSRSGVFIGACNDDYHCM
                     QVERPETGDAFSATGVAASVLSGRLSYLFNLQGPSLVVDTACSSSLVSLHLACQSLRN
                     RECNMALAGGVNLILSPQSVLLVSKLQALSPDGRSKAFDASANGFTRGEGCGIVVLKR
                     LSDALADGDHILATIRGSAINQDGKSTGLTTPNVLSQQALIRQALESSGLKAEQVSYV
                     EAHGTGTPLGDPIEAEALREVYGVPRPDGSVCGIGSVKTNVGHLESAAGIAGIMKVVL
                     AMRHQSIPPHLHLKQVNPRIQLEGSALTIPTRLTPWKASGQPRRAGVSSFGISGTNAH
                     VIIEEAPAPVAKDALPQRPIVLPLSARSREALVALAERYAERLGREGGGSLSDVCYTA
                     SAHRTHHTHRLAVLADSREHTAERLRAFARGEQPSGAWSGRKAFGGRRKIVFVYPGQG
                     AQRAGMGRALMESEPVFRDALVKVDKALQPHLGWSVLDEIAADEQRSRLAEIDVNQPV
                     LFAVEVALTALWRSWGIEPDAVMGHSMGEVVAAHVAGALSLEDAVLIICHRSKLMRTL
                     GGRGATMAMVELSVEEAEEAIRPWNGKIWLGGLNGPRSQVLSGEPEAVKEAVEALGKR
                     GVFTRWVKMDVPSHTPLVEVISRELEGRIARITPRAADIPLCSTVTGEFLDGAKLDAA
                     HWGNNLRNPVRFAAAAQHLCASDHGTFVEVSPHPVLLPAVEQTMSAGNTEGVVLASLR
                     RNEDERHTMLSSLGALYSSGHRVAWNRVYPGKGQLVPLPTYPWQRERFWMEPAVRGVR
                     ASRTSSGHPLIGERIDLSAQGEVSRLWQVQLGPKHFPLLAEHRVQGVPALPAAAMADM
                     ALSAAKQTLGEPVSLEDVSFSALLSFPEDEERVLQLWLSTERPGASAFRLFSAPVMRE
                     TPSDELTWALHAEGSLRHGDASVVEKAAPVALDELRARCGEEASVEAHYQTLKEIGLE
                     YGPSFRVVRELYRAKGEALSRVQLEGELTAEEERYSLHPALLDGCFQVLAAALEGLDS
                     EGGVYVPVALSRLRVLVRPEAALWSHVRIVSGAEAGSNTIECDVTLLNDQGRVVAEVR
                     RLVCKKIAAASAPADEVDRWIYSAQWEPQQRQARAETGSSRTAGRWLVFADARGIGQS
                     LASGLAERAGGCVVVEAGPSYERLEADRYRINPSKLEDFQALLKEAFSDQSPCKGMVH
                     LYGLDVASNVTEPESFERALSLSSTSVVYLLQAWAQRGWRDTPRLWLVTTGVQVIASG
                     EQPGSPAAAPLVGLAAVITHEHPELRCAHVDLSAQPSPEEIASLTEELLGDGAEDRIA
                     LRGAVRHVARLGRGISGRKAAEPREFTADGTYLITGGLGGLGLEVAKWFIQRGARNLA
                     LTGRGAPSEEAQKVLGELEANGAEVRVYRADVTSAADVARLLGSIDSELPPLRGVVHA
                     AGVIADGVLTQLDEPRLRAVLAPKVQGAWNLHVQTQGRPLELFVMFSSAASMMGSPGQ
                     GNYAAANAFMDVLAHHRRAAGLPGLSVNWGPWADVGLAAVATNRGDRLALRGVGNMPP
                     AQALEALGRLLVQEAAQVAVVPLDLRQWREFYLSAAQSPFLSVLMQEKVSTPASRKGS
                     ARELLAAAEPEQRKGILECYLREQLGRILRMEPSRIEPDQPFGDLGMDSLTGLELRNR
                     IESGLKLTLSATLIYAYPTVQALTTYLYGKLEPVLEAEARPTVRPTQTEPRVSVGVAP

  Query Match          61.3%;  Score 55.8;  DB 1;  Length 49736;
  Best Local Similarity  75.8%;  Pred. No. 0.0004;
  Matches   69;  Conservative    0;  Mismatches   22;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||| ||||    ||| || || | ||||  ||||  ||| || |||||||| |||||||
Db   29046 CCGATGCGCTGGCGGACGGTGATCACATCCTGGCGACGATTCGCGGCTCGGCGATCAACC 28987

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | ||||||  |||||| ||| ||||| ||||
Db   28986 AGGATGGCAAGTCGACAGGGCTGACCACGCC 28956


RESULT 13
AB070940
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
     gene            3213. .3971
                     /gene="olmC"
     CDS             3213. .3971
                     /gene="olmC"
                     /codon_start=1
```

```
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
gene                 4643. .7438
                     /gene="olmRI"
CDS                  4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
gene                 7658. .10486
                     /gene="olmRII"
CDS                  7658. .10486
                     /gene="olmRII"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69190.1"
                     /db_xref="GI:15823973"
                     /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                     AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                     RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                     QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                     LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                     DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                     HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                     EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                     GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                     VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                     EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                     VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                     GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                     PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                     LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                     LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                     ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS                  10933. .11271
                     /codon_start=1
                     /transl_table=11
                     /product="P450-like protein"
                     /protein_id="BAB69191.1"
                     /db_xref="GI:15823974"
                     /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                     SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                     PCPALPPDRP"
gene                 11458. .29898
                     /gene="olmA1"
CDS                  11458. .29898
                     /gene="olmA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69192.1"
                     /db_xref="GI:15823975"
                     /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                     VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                     DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                     AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                     VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                     AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                     LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                     GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                     HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                     ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                     AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                     TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                     AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                     LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                     TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                     WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                     DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                     LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                     DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                     QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                     PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                     GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                     SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                     LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                     PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                     DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                     EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                     EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                     SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                     AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                     AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                     GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                     PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                     VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                     LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                     GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                     GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                     AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                     TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                     LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                     DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

```
  Query Match           59.6%;  Score 54.2;  DB 1;  Length 104326;
  Best Local Similarity 74.7%;  Pred. No. 0.00085;
  Matches   68;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       1 CCGACGCGCAACGGGATGGCGACCGCATCTGGGCGCTGATCCGGGGCTCGGCCATCAACC 60
           |||||||||  ||||| ||  ||| | ||  |||| ||||||| || |||||  | ||||
Db   57234 CCGACGCGCGCCGGGACGGTCACCCCGTCCTGGCGGTGATCCGCGGGTCGGCGGTGAACC 57293

Qy      61 ATGATGGCCGGTCGACCGGGTTGACCGCGCC 91
           | || |||  ||||| |||  ||| ||||||
Db   57294 AGGACGGCGCGTCGAACGGTCTGAGCGCGCC 57324


RESULT 14
AX024383
LOCUS       AX024383               10910 bp    DNA     linear   BCT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024383
VERSION     AX024383.1  GI:10184587
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
```

```
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match             57.4%;  Score 52.2;  DB 1;  Length 10910;
  Best Local Similarity   74.2%;  Pred. No. 0.0038;
  Matches   66;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      3 GACGCGCAACGGGATGGCGACCGCATCTGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
          |||||||| || ||  |||||| ||||  |||| || || || || |||| ||||||||
Db   3000 GACGCGCAGCGCGACCGCGACCCCATCCTGCGCGGTGGTCAGGAGCACGGCGATCAACCAC 3059

Qy     63 GATGGCCGGTCGACCGGGTTGACCGCGCC 91
          ||||||| |   | |||| | || | |||
Db   3060 GATGGCCCGAGCAGCGGGCTCACGGTGCC 3088


RESULT 15
AX024276
LOCUS       AX024276               10910 bp    DNA     linear   PAT 15-SEP-2000
DEFINITION  Sequence 81 from Patent DE19846493.
ACCESSION   AX024276
VERSION     AX024276.1  GI:10184550
KEYWORDS    .
SOURCE      Polyangium cellulosum.
  ORGANISM  Polyangium cellulosum
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Sorangineae; Polyangiaceae; Polyangium.
REFERENCE   1  (bases 1 to 10910)
  AUTHORS   Beyer,S. and Mueller,R.J.
  JOURNAL   Patent: DE 19846493-A 81 13-APR-2000;
            BIOTECHNOLOG FORSCHUNG GMBH (DE)
FEATURES             Location/Qualifiers
     source          1. .10910
                     /organism="Polyangium cellulosum"
                     /db_xref="taxon:56"
BASE COUNT     1852 a   3124 c   3613 g   2321 t
ORIGIN

  Query Match             57.4%;  Score 52.2;  DB 6;  Length 10910;
  Best Local Similarity   74.2%;  Pred. No. 0.0038;
  Matches   66;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      3 GACGCGCAACGGGATGGCGACCGCATCTGGCGCTGATCCGGGGCTCGGCCATCAACCAT 62
          |||||||| || ||  |||||| ||||  |||| || || || || |||| ||||||||
Db   3000 GACGCGCAGCGCGACCGCGACCCCATCCTGCGCGGTGGTCAGGAGCACGGCGATCAACCAC 3059

Qy     63 GATGGCCGGTCGACCGGGTTGACCGCGCC 91
          ||||||| |   | |||| | || | |||
Db   3060 GATGGCCCGAGCAGCGGGCTCACGGTGCC 3088


Search completed: November  5, 2002, 12:05:23
Job time : 561.361 secs
```